

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:56 ; Search time 42 Seconds
(without alignments)
452.323 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGKLDIPISDFGRMGI.....QAINQLDIPPLMISYLKHL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/2/1aa/5A-COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5B-COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A-COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B-COMB.pep: *
5: /cgn2_6/prodata/2/1aa/6CTUS-COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480	20.5	350	4	US-09-302-769-25
2	293	12.5	458	4	US-09-811-469-10
3	293	12.5	458	4	US-09-811-469-11
4	264	11.3	621	3	US-09-121-964-8
5	264	11.3	621	4	US-09-155-078-2
6	219	9.3	1166	4	US-09-350-982C-5
7	209	8.9	949	3	US-09-196-387-10
8	209	8.9	949	4	US-09-841-835-10
9	208.5	8.9	1327	3	US-09-196-387-2
10	208.5	8.9	1327	4	US-09-841-835-2
11	202.5	8.6	1839	2	US-09-172-977-4
12	202	8.6	1088	3	US-09-082-059-2
13	201.5	8.6	303	2	US-09-031-485-23
14	201.5	8.6	303	2	US-08-847-429A-23
15	201.5	8.6	303	3	US-09-065-474-23
16	201.5	8.6	303	4	US-09-557-034-23
17	201.5	8.6	352	3	US-09-065-474-139
18	201.5	8.6	352	4	US-09-557-034-139
19	201.5	8.6	1745	2	US-09-031-485-33
20	201.5	8.6	1745	2	US-08-847-429A-33
21	201.5	8.6	1745	3	US-09-065-474-33
22	201.5	8.6	1745	3	US-09-065-474-33
23	198	8.5	843	2	US-09-172-977-3
24	195	8.3	1619	4	US-09-392-812A-4
25	195	8.3	1704	4	US-09-392-812A-2
26	194.5	8.3	302	2	US-09-031-485-38
27	194.5	8.3	302	2	US-08-847-429A-38

28	194.5	8.3	302	3	US-09-065-474-38	Sequence 38, Appl
29	194.5	8.3	302	4	US-09-557-034-38	Sequence 38, Appl
30	191.5	8.2	673	3	US-09-196-387-8	Sequence 8, Appl
31	191.5	8.2	673	4	US-09-841-835-8	Sequence 8, Appl
32	176.5	7.5	1709	4	US-09-392-812A-6	Sequence 6, Appl
33	168.5	7.2	679	2	US-08-462-481-4	Sequence 6, Appl
34	168.5	7.2	679	2	US-08-436-771-6	Sequence 6, Appl
35	168.5	7.2	679	2	US-08-434-998-6	Sequence 6, Appl
36	168.5	7.2	679	2	US-08-487-797-6	Sequence 6, Appl
37	168.5	7.2	679	2	US-08-701-005A-4	Sequence 4, Appl
38	168.5	7.2	679	2	US-08-479-895-4	Sequence 4, Appl
39	168.5	7.2	679	3	US-08-943-956A-4	Sequence 4, Appl
40	168.5	7.2	679	5	PCT-US95-02058-6	Sequence 6, Appl
41	164	7.0	191	2	US-09-031-485-20	Sequence 20, Appl
42	164	7.0	191	2	US-08-847-429A-20	Sequence 20, Appl
43	164	7.0	191	3	US-09-065-474-20	Sequence 20, Appl
44	164	7.0	191	4	US-09-557-034-28	Sequence 20, Appl
45	161.5	6.9	348	2	US-09-031-485-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-302-769-25
; Sequence 25, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 10976Z
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (167)
; OTHER INFORMATION: Xaa is unsure
US-09-302-769-25

Query Match	20.5%	Score 480;	DB 4;	Length 350;
Best Local Similarity	38.3%	Pred. No. 8.4e-45;		
Matches 114;	Conservative 58;	Mismatches 104;	Indels 22;	Gaps 8;
QY	170	LNQAVYDNDSTLDQLLRQERYKRFINSRSGW---	GVPGTPLRLAASYGHLSCLOVLLAH	226
Db	56	LHDAAYVGLQTLRLNLQEEYSRINEKSVCCGWLCTPLRIATAAGHNCVDFLRK	115	
QY	227	GADVDSLVKQATPLFTAVSHGHLDVCVLLFAGASPGSIYNNCSPLVTAARDGAVAIL	286	
Db	116	GAEVDLVKQGTALYVAVNGHLESTLELLFAGADPNRHRSTPPVYHXXVGRDDIL	175	
QY	287	QELLDHGAENVKAKL-----PVWASNTAS---	CSGPLYLAAYGHLCDFRLLHGGADP	338
Db	176	KALTRYGADVDDVNNHLSNDRTPPFRRLLTSLVVC--	PLYISAAYHNLQCFRLLQAGANP	233
QY	339	DYNCCTDQGLLARVPR--PRTLEICLHNCPEYIQLIDFGANIYV---	PSLSLDLTSG	393

```
Db 234 DFNCGPVTQEFYRSGPCWGMDVLRHGCFAAFVLLVEFGANLNLVKWESLGPARGR 293
Qy 394 ---DDKGIALLOARPTPSLLSQVRLVVRALCOAQOQAINQLDIPWMLISYLKHQ 448
Db 294 RKMDPALQVFKEARIPRTLLSLCRVAVRALKGY-RLHLVPSLPLPDPIKKFLLYE 350

RESULT 2
US-09-811-469-10
; Sequence 10, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; AND USES THEREOF
; FILE REFERENCE: CL001171
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Human
US-09-811-469-10

Query Match 12.5%; Score 293; DB 4; Length 458;
Best Local Similarity 27.7%; Pred. No. 1.1e-23;
Matches 86; Conservative 49; Mismatches 91; Indels 84; Gaps 13;

Qy 3 ESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSK 62
Db 129 ENGWKLYDPLEYRRQGGIPNESWRITKINERYELCDTPALLVVPANIPDEELKRVASPR 188
Qy 63 SKERPVLISYKKNNAIARCSCQPSLGSFY-TRCVDELLELLEAISQTNPGSQFMVYVDTR 121
Db 189 SRGRIPVLSWIHPESQATITRCSQPMVGVSGKSKDEKYLQAIMDSNAQSHKIFIDAR 248
Qy 122 PKI-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 249 PSVNAVANKAGGYESDAYQNAELVF---LDIHNHVMRESLRKLEIVYVNIETHW 305
Qy 160 ----EDT-----DTEEQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHLEHIKILAGALRIADKVESGKTSVVVHCSGDGWDRTAQLTSLAMLMDGY 365
Qy 193 RFINS-----RSGWGVPGTPLRLAASVGHSLCQLVLLAHG---ADVDSLKVKAQTPFLT 243
Db 366 RTIRGFEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
Qy 244 AVSHGLDCV 253
Db 408 Q----FIDCV 413

Query Match 12.5%; Score 293; DB 4; Length 458;
Best Local Similarity 27.7%; Pred. No. 1.1e-23;
Matches 86; Conservative 49; Mismatches 91; Indels 84; Gaps 13;

Qy 3 ESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSK 62
Db 129 ENGWKLYDPLEYRRQGGIPNESWRITKINERYELCDTPALLVVPANIPDEELKRVASPR 188
Qy 63 SKERPVLISYKKNNAIARCSCQPSLGSFY-TRCVDELLELLEAISQTNPGSQFMVYVDTR 121
Db 189 SRGRIPVLSWIHPESQATITRCSQPMVGVSGKSKDEKYLQAIMDSNAQSHKIFIDAR 248
Qy 122 PKI-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 249 PSVNAVANKAGGYESDAYQNAELVF---LDIHNHVMRESLRKLEIVYVNIETHW 305
Qy 160 ----EDT-----DTEEQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHLEHIKILAGALRIADKVESGKTSVVVHCSGDGWDRTAQLTSLAMLMDGY 365
Qy 193 RFINS-----RSGWGVPGTPLRLAASVGHSLCQLVLLAHG---ADVDSLKVKAQTPFLT 243
Db 366 RTIRGFEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
Qy 244 AVSHGLDCV 253
Db 408 Q----FIDCV 413

Query Match 11.3%; Score 264; DB 3; Length 621;
Best Local Similarity 26.7%; Pred. No. 3.1e-20;
Matches 80; Conservative 46; Mismatches 106; Indels 69; Gaps 12;

Qy 5 GWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSKFSK 64
Db 181 GMTVYNPVEYRRQGLPNHHRITTFINKCYELCDTPALLVVPYRASDDDLRRVATFRSR 240
Qy 65 ERVPVLSYKKNNAIARCSCQPSLGSFY-TRCVDELLELLEAISQTNPGSQFMVYVDTRPK 123
Db 241 NRIPVLSWIHPENKTVIVRCQPSLPGMSGKRNKDKDEKYLVDIRETNKQISKLTIYDARPS 300
Qy 124 I-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 301 VNAVANKATGGGYESDDAYHNAELFF---LDIHNHVMRESLRKLEIVYVNIETHW 357
Qy 160 --EDTTEE--KQALNQAVYDNDSTYLDQLLRQERYKRFINSRSGVPGTFLRLA--- 211
Db 358 SLESTHLEHIKILVTGAIQVADK-----VSSGSSVLVHCSGDGWDRTAQLTSLAMLML 411
Qy 212 -----ASYGHLSCLQVLLAHGADVDSLKVKAQTPFLTAVSHGLDCV 253
Db 412 DSFYRSIGFEILVQKEWISFGHKFASRI--CHG-DKNHTDAD-RSPIFLQ----FIDCV 463
```

```
Db 234 DFNCGPVTQEFYRSGPCWGMDVLRHGCFAAFVLLVEFGANLNLVKWESLGPARGR 293
Qy 394 ---DDKGIALLOARPTPSLLSQVRLVVRALCOAQOQAINQLDIPWMLISYLKHQ 448
Db 294 RKMDPALQVFKEARIPRTLLSLCRVAVRALKGY-RLHLVPSLPLPDPIKKFLLYE 350

RESULT 2
US-09-811-469-10
; Sequence 10, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; AND USES THEREOF
; FILE REFERENCE: CL001171
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Human
US-09-811-469-10

Query Match 12.5%; Score 293; DB 4; Length 458;
Best Local Similarity 27.7%; Pred. No. 1.1e-23;
Matches 86; Conservative 49; Mismatches 91; Indels 84; Gaps 13;

Qy 3 ESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSK 62
Db 129 ENGWKLYDPLEYRRQGGIPNESWRITKINERYELCDTPALLVVPANIPDEELKRVASPR 188
Qy 63 SKERPVLISYKKNNAIARCSCQPSLGSFY-TRCVDELLELLEAISQTNPGSQFMVYVDTR 121
Db 189 SRGRIPVLSWIHPESQATITRCSQPMVGVSGKSKDEKYLQAIMDSNAQSHKIFIDAR 248
Qy 122 PKI-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 249 PSVNAVANKAGGYESDAYQNAELVF---LDIHNHVMRESLRKLEIVYVNIETHW 305
Qy 160 ----EDT-----DTEEQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHLEHIKILAGALRIADKVESGKTSVVVHCSGDGWDRTAQLTSLAMLMDGY 365
Qy 193 RFINS-----RSGWGVPGTPLRLAASVGHSLCQLVLLAHG---ADVDSLKVKAQTPFLT 243
Db 366 RTIRGFEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
Qy 244 AVSHGLDCV 253
Db 408 Q----FIDCV 413

Query Match 11.3%; Score 264; DB 3; Length 621;
Best Local Similarity 26.7%; Pred. No. 3.1e-20;
Matches 80; Conservative 46; Mismatches 106; Indels 69; Gaps 12;

Qy 5 GWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSKFSK 64
Db 181 GMTVYNPVEYRRQGLPNHHRITTFINKCYELCDTPALLVVPYRASDDDLRRVATFRSR 240
Qy 65 ERVPVLSYKKNNAIARCSCQPSLGSFY-TRCVDELLELLEAISQTNPGSQFMVYVDTRPK 123
Db 241 NRIPVLSWIHPENKTVIVRCQPSLPGMSGKRNKDKDEKYLVDIRETNKQISKLTIYDARPS 300
Qy 124 I-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 301 VNAVANKATGGGYESDDAYHNAELFF---LDIHNHVMRESLRKLEIVYVNIETHW 357
Qy 160 --EDTTEE--KQALNQAVYDNDSTYLDQLLRQERYKRFINSRSGVPGTFLRLA--- 211
Db 358 SLESTHLEHIKILVTGAIQVADK-----VSSGSSVLVHCSGDGWDRTAQLTSLAMLML 411
Qy 212 -----ASYGHLSCLQVLLAHGADVDSLKVKAQTPFLTAVSHGLDCV 253
Db 412 DSFYRSIGFEILVQKEWISFGHKFASRI--CHG-DKNHTDAD-RSPIFLQ----FIDCV 463
```

RESULT 5

US-09-155-078-2
Sequence 2, Application US/09155078
Patent No. 6312688
GENERAL INFORMATION:
APPLICANT: FOUSTKA, Annemarie
APPLICANT: KIOSCHIS, Petra
APPLICANT: LAPORTE, Jocelyn
APPLICANT: HU, Ling Jia
APPLICANT: MANDEL, Jean Louis
APPLICANT: DAHL, Niklas
TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
FILE REFERENCE: 012627-010
CURRENT APPLICATION NUMBER: US/09/155,078
CURRENT FILING DATE: 1999-03-02
EARLIER APPLICATION NUMBER: PCT/DE97/00592
EARLIER FILING DATE: 1997-03-21
EARLIER APPLICATION NUMBER: DE 196 11 234.6
EARLIER FILING DATE: 1996-03-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 621
TYPE: PRT
ORGANISM: German-type microorganism & cell cul.
US-09-155-078-2

Query Match 11.3%; Score 264; DB 4; Length 621;
Best Local Similarity 26.7%; Pred. No. 3.1e-20;
Matches 80; Conservative 46; Mismatches 106; Indels 68; Gaps 12;

QY 5 GWKLIDPISDFGRMGIPNRRNTITDANRYICSTYPPEIIVPKSVTLGTWVGSSKRSK 64
DB 181 GTVYNVEEYRQLPNHHWRITINKCYELCTYPALLVPYRASDDDLRRVATFRSR 240
QY 65 ERPVLSYLKENNAACRCOSPLSGFY-TRCVDDELLEAISOTFPGSQPMYYVDTRPK 123
DB 241 NRIPVLSWHPENKTIVRCSPQVMGSKRNKDKEYLDVIRETNKQISKLIYDARPS 300
QY 124 I-----WHFLVLMRVILQLAKNLN--DIKIFSLQPDKEE----- 159
DB 301 VNAVANKATGGYESDDAYHNAELFF---LOITHNVMRRESLKVKYOIVYFNVEESHWLS 357
QY 160 --EDTTEE--KOALNQAVYDNDSVTLDQLLRQERYKFIRNSRGWGVPGLRLA--- 211
DB 358 SLETHWLEHIKLVLTGAQVADK-----VSSGKSXSVLVHCSDGNWRTAQLTSLAMLML 411
QY 212 -----ASYGHLSCLQVLLAHGADVDSLQVKAQTPLFTAVSHGHLDVCV 253
DB 412 DSFYRSIEGFELVQKEWISFGHKPASRI--GHG-DKNHTDAD-RSFIPLQ----FDICV 463

RESULT 6

US-09-350-982C-5
Sequence 5, Application US/09350982C
Patent No. 6455290
GENERAL INFORMATION:
APPLICANT: Bertheisen, Jens
APPLICANT: Toma, Salvatore
APPLICANT: Isacchi, Antonella
TITLE OF INVENTION: Tankyrase Homolog Protein(THP), Nucleic Acids, and Methods Relat
FILE REFERENCE: PHRM-0043
CURRENT APPLICATION NUMBER: US/09/350,982C
CURRENT FILING DATE: 1999-07-09
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 1166
TYPE: PRT
ORGANISM: Artificial

FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (1102)...(1102)
OTHER INFORMATION: n is any nucleic acid
NAME/KEY: misc feature
LOCATION: (2650)..(2650)
OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-5

Query Match 9.3%; Score 219; DB 4; Length 1166;
Best Local Similarity 30.3%; Pred. No. 9.7e-15;
Matches 74; Conservative 26; Mismatches 78; Indels 66; Gaps 7;

QY 189 ERYKRF-----INSRSGMVGPGTPLRLAAASYGHLSCLQVLLAHGADVDSLQVKAQTPLFT 243
DB 38 ERVKELVTPKEKVNRSRDTAGRKSTPLHLAAGFGKRDVVGYELLQNGANNVQARDGGLIPLHN 97
QY 244 AVSHGHLCVRLLEAGASPGGSYNNCSVPLTAARDGAVAILQELLDHGAEANVKAQLP 303
DB 98 ACSFGHAEEVNNLLRHGADPNARNWNYPPLHEAAIKGDKIDVICVILLQHGAETPIRTNDG 157
QY 304 VWASNIA-----SC-----SGPLYL 318
DB 158 RTALDLADPSAKAVITGEYKKDELLESARSNGEEKMMALLTPLNVNCHASDGKRKSTPLHL 217
QY 319 AAVYGHLCDFRLLIHGADPDYNCCTDQGLLARVPRPTLLEICLIHMC---EPEYIOULL 375
DB 218 AAGYNRVKIQQILLQHGA--DVHAKDKGDL--VP-----LHNACSXYHYEVTELLV 264
QY 376 DFGA 379
DB 265 KHGA 268

RESULT 7

US-09-196-387-10
Sequence 10, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5900
TELEFAX: 201-343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10

Query Match 8.9%; Score 209; DB 3; Length 949;

Best Local Similarity 27.4%; Pred. No. 9e-14;
Matches 93; Conservative 30; Mismatches 122; Indels 94; Gaps 11;

QY	139	AKNLMIDTIKIFSLLOPDKEEDTTEKQALNQAVYDNDSTLQDQLLRQERYKRFINSR	198
DB	628	AQMGNEAVQOILSESTPIR-TSDVD-----YRLLEASKAGDLETVKOLCSSQN-----VNCR	678
QY	199	SCMGVPGTPLRLAASVGHLSCLQVLLAHGADVDSLQVKAQTPLFTAVSHGHLDQVVRVLE	258
DB	679	DLEGRHSTPLHFAAGYNRVSVVEYLLHHGADVHAKDKGGLVPLHNACSYGHVEVAELLVR	738
QY	259	AGASPGGSYNNCSVPLTAARDGAVAILQELLDHGAENVK-----	299
DB	739	HGASVNVADLWKFTPLHEAAAKGYEICKLLKHGADPTKKNRDGNTPLDLVKEGDTDIQ	798
QY	300	-----AKLPVWA-----SNIASC-----SGPLYLAAYVGHLDQVRLLL	332
DB	799	DLKGDAAALLDAKKGCLARVQKLTCPENI-NCRTQGRNSTPLHLAAGYNNLEVAEYLL	857
QY	333	LHGAD-----PDYNCCTDQ-GLLARVPRPTLLLEICLHNCPEYIQLLIDFGANI	381
DB	858	EHGADVNAQDKGLIPLHNAASYGGCLARVQK-----LCTPENINCRDTQGRNS	906
QY	382	-----YLPSSLDLTSDQDKGIALLLQA	404
DB	907	TPHLAAGYNNLEVAEYLLLEHGADVNAQDKGGLIPLHNA	945

RESULT 8

US-09-841-835-10

Sequence 10, Application US/09841835

Patent No. 6506587

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,835

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/196,387

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 949 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-841-835-10

Query Match

Best Local Similarity 8.9%; Score 209; DB 4; Length 949;

Matches 93; Conservative 30; Mismatches 122; Indels 94; Gaps 11;

QY	139	AKNLMIDTIKIFSLLOPDKEEDTTEKQALNQAVYDNDSTLQDQLLRQERYKRFINSR	198
DB	628	AQMGNEAVQOILSESTPIR-TSDVD-----YRLLEASKAGDLETVKOLCSSQN-----VNCR	678
QY	199	SCMGVPGTPLRLAASVGHLSCLQVLLAHGADVDSLQVKAQTPLFTAVSHGHLDQVVRVLE	258
DB	679	DLEGRHSTPLHFAAGYNRVSVVEYLLHHGADVHAKDKGGLVPLHNACSYGHVEVAELLVR	738
QY	259	AGASPGGSYNNCSVPLTAARDGAVAILQELLDHGAENVK-----	299
DB	739	HGASVNVADLWKFTPLHEAAAKGYEICKLLKHGADPTKKNRDGNTPLDLVKEGDTDIQ	798
QY	300	-----AKLPVWA-----SNIASC-----SGPLYLAAYVGHLDQVRLLL	332
DB	799	DLKGDAAALLDAKKGCLARVQKLTCPENI-NCRTQGRNSTPLHLAAGYNNLEVAEYLL	857
QY	333	LHGAD-----PDYNCCTDQ-GLLARVPRPTLLLEICLHNCPEYIQLLIDFGANI	381
DB	858	EHGADVNAQDKGLIPLHNAASYGGCLARVQK-----LCTPENINCRDTQGRNS	906
QY	382	-----YLPSSLDLTSDQDKGIALLLQA	404
DB	907	TPHLAAGYNNLEVAEYLLLEHGADVNAQDKGGLIPLHNA	945

RESULT 9

US-09-196-387-2

Sequence 2, Application US/09196387

Patent No. 627613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

US-09-172-977-4

Query Match 8.6%; Score 202.5; DB 2; Length 1839;
Best Local Similarity 26.8%; Pred. No. 1.5e-12;
Matches 63; Conservative 31; Mismatches 84; Indels 57; Gaps 4;
QY 206 TPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHDCVRVLLVLEAGASPGG 265
DB 565 TPLHVAAGYSLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLEKASPHA 624
QY 266 SIYNCSPLVTAARDGAVAILQELLDHGAENVKAKLPWASNIASCSGPLYLAAYVGH 325
DB 625 TAKNGYTPHIAAKNQMQASTLLNYGAETNIVTKGV-----TPLHLASQEGHT 675
QY 326 DCFRLLLHGAADPDYNTDQGLLARVPRPRTLLLEICLHNCEPEYIQLLIDFGANIYLP 385
DB 676 D-----MTLLDQKGANIHMT 692
QY 386 ----LSLDTSDDK-GIALLQARATPRSLLSQVRLVVRALCOAGQFQAINOL 435
DB 693 KSLGTLHLAAQEDKVNADILTKHADQDAHTKLGVTPLIVACHYGNVKNVNF 747

RESULT 12

US-09-082-059-2
; Sequence 2, Application US/09082059A
; Patent No. 6225086

; GENERAL INFORMATION:

; APPLICANT: Morrow, Jon S.

; APPLICANT: Devarajan, Prasad

; TITLE OF INVENTION: No. 6225086el Ankyrin Proteins and a Method for Their Identification

; FILE REFERENCE: 44574-5002-US

; CURRENT APPLICATION NUMBER: US/09/082,059A

; CURRENT FILING DATE: 1998-05-21

; EARLIER APPLICATION NUMBER: 60/047356

; EARLIER FILING DATE: 1997-05-21

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 1088

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-082-059-2

Query Match 8.6%; Score 202; DB 3; Length 1088;
Best Local Similarity 27.3%; Pred. No. 6.9e-13;
Matches 71; Conservative 34; Mismatches 89; Indels 66; Gaps 8;
QY 206 TPLRLAASVGHLSCLQVLLAHGADVDSLDVKA----- 237
DB 56 TPLHVAAGYSLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLEKASPHA 624
QY 238 ----QTPLFTAVSHGHDCVRVLLVLEAGASPGGSIYNCSPLVTAARDGAVAILQELLDH 292
DB 116 KAKDDQTPHLSARLKADIVQQLQOGASPNAAATSGYTPHLSAREGHEDVAAPFLDH 175
QY 293 GAENVKAKLPWASNIASCSG--PLYLAAYVGHLCFRLLILHGAADPDYNTDQGLLAR 350
DB 176 GASLSITTK-----KGTTPHVAAGYSLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLEKASPHA 624
QY 351 VPRPRTLLLEICLHNCEPEYIQLLIDFGA-----NLYLPSLSLDTSDDKGIAL-- 401
DB 221 ----TPLHVAHYD-NQKVALLEKASPHAAGKNGYTPHIAAKNQMDIATLLLEY 274
QY 402 -LOARATPRSLLSQVRLVVR 420
DB 275 GADANAVTRQGIASVHLAAQ 294

RESULT 13

US-09-031-485-23

; Sequence 23, Application US/09031485

; Patent No. 5824306

; GENERAL INFORMATION:

; APPLICANT: Tang, Liang

; APPLICANT: Blehm, E. Scot

; TITLE OF INVENTION: DIOFILARIA AND BRUGIA ANKYRIN

; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carol Talkington Verser, Ph.D.

; ADDRESSEE: Heeka Corporation

; STREET: 1825 Sharp Point Drive

; CITY: Fort Collins

; STATE: Colorado

; COUNTRY: USA

; ZIP: 80525

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Windows 95

; SOFTWARE: WordPerfect for Windows, Version 7.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/031,485

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/847,429

; FILING DATE: 24-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Verser, Carol Talkington

; REGISTRATION NUMBER: 37,459

; REFERENCE/DOCKET NUMBER: HW-5

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 970/493-7272

; TELEFAX: 970/494-9505

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 303 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-031-485-23

Query Match 8.6%; Score 201.5; DB 2; Length 303;
Best Local Similarity 26.9%; Pred. No. 9e-14;
Matches 66; Conservative 36; Mismatches 78; Indels 65; Gaps 6;
QY 206 TPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHDCVRVLLVLEAGASPGG 265
DB 109 TPLHVAAGYSLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLEKASPHA 168
QY 266 SIYNCSPLVTAARDGAVAILQELLDHGAENVKAKLPWASNIASCSG--PLYLAAYVGH 323
DB 169 ATRDYTPHIAAKNGYSLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLEKASPHA 217
QY 324 HLDCEFRLLILHGAADPDYNTDQGLLARVPRPRTLLLEICLHNCEPEYIQLLIDFGANIY 383
DB 218 NLPVAKSLLETCPTVDIEGKQ-----VTPHVAHYDNQKVALLEKASPHA 250
QY 384 PSLSLDTSDDKGIAL--ARATPRSLLSQVRLVVRALCOAGQFQAINOL 439
DB 251 ----NDKVALLEKASPHAAGKNGYTPHIAAK-----NQMDIAS 289
QY 440 MLISY 444
DB 290 TLLHY 294

RESULT 14

US-08-847-429A-23

; Sequence 23, Application US/08847429A

; Patent No. 5827692

; GENERAL INFORMATION:

; APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: US\$ THEREOF
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/47,429A
FILING DATE: 24-APR-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-847-429A-23

Query Match 8.6%; Score 201.5; DB 2; Length 303;
Best Local Similarity 26.9%; Pred. No. 9e-14;
Matches 66; Conservative 36; Mismatches 78; Indels 65; Gaps 6;

```
QY 206 TPLRLAASYGHLSCQLVLLAHGADVSLDVKAQTPLFTAVSHGHLDVCRVLLLEAGASPGG 265
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 TPLHLAARANQTDIVRVLVRNGAQVDAARELQTPHLIASRLGNTDIVILLQANASPN 168
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 SIYNNCSPLVTAARDGAVAILQELLDHGAENNVKALPWASNIASCSG--PLYLAAYVG 323
: : : : : : : : : : : : : : : : : : : : : : : :
Db 169 ATRDLYTPLHIAAKEGQEEVAAILMDHGTDKTLTK-----KGFTPLHLAAKYG 217
: : : : : : : : : : : : : : : : : : : : : : : :
QY 324 HLDGCFRLLLHGGADPDYNTDQGLLARVPRPTLLEICLHNCEPEYIQLLIDFGANIYL 383
: : : : : : : : : : : : : : : : : : : : : : : :
Db 218 NLPVAKSLLERGTPVDIEGKNQ-----VTPHLVAAHYN----- 250
: : : : : : : : : : : : : : : : : : : : : : : :
QY 384 PSLSLDLTSQDDKGIALLLQ-----ARATPRSLLSQVRLVRRALCQAGOPAINQLDIPP 439
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 -----NDKVALLLENGASAHAAKNGYTPHLHIAAK-----NQMDIAS 289
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 440 MLISY 444
: : : : : : : : : : : : : : : : : : : : : : : :
Db 290 TLLHY 294
```

RESULT 15

US-09-065-474-23

; Sequence 23, Application US/09065474
; Patent No. 6063599

GENERAL INFORMATION:

APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scot

TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

TITLE OF INVENTION: US\$ THEREOF

NUMBER OF SEQUENCES: 171

CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/065,474
FILING DATE: 24-APR-1998
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-065-474-23

Query Match 8.6%; Score 201.5; DB 3; Length 303;
Best Local Similarity 26.9%; Pred. No. 9e-14;
Matches 66; Conservative 36; Mismatches 78; Indels 65; Gaps 6;

```
QY 206 TPLRLAASYGHLSCQLVLLAHGADVSLDVKAQTPLFTAVSHGHLDVCRVLLLEAGASPGG 265
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 TPLHLAARANQTDIVRVLVRNGAQVDAARELQTPHLIASRLGNTDIVILLQANASPN 168
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 SIYNNCSPLVTAARDGAVAILQELLDHGAENNVKALPWASNIASCSG--PLYLAAYVG 323
: : : : : : : : : : : : : : : : : : : : : : : :
Db 169 ATRDLYTPLHIAAKEGQEEVAAILMDHGTDKTLTK-----KGFTPLHLAAKYG 217
: : : : : : : : : : : : : : : : : : : : : : : :
QY 324 HLDGCFRLLLHGGADPDYNTDQGLLARVPRPTLLEICLHNCEPEYIQLLIDFGANIYL 383
: : : : : : : : : : : : : : : : : : : : : : : :
Db 218 NLPVAKSLLERGTPVDIEGKNQ-----VTPHLVAAHYN----- 250
: : : : : : : : : : : : : : : : : : : : : : : :
QY 384 PSLSLDLTSQDDKGIALLLQ-----ARATPRSLLSQVRLVRRALCQAGOPAINQLDIPP 439
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 -----NDKVALLLENGASAHAAKNGYTPHLHIAAK-----NQMDIAS 289
||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 440 MLISY 444
: : : : : : : : : : : : : : : : : : : : : : : :
Db 290 TLLHY 294
```

Search completed: December 1, 2003, 07:38:05

Job time : 44 secs

THIS PAGE BLANK (USPTO)

Db 347 GGGGCGAGGTGGACCTGGTGGATGTCAGGGGCGAGACTGCCCTGTATGTGGCTGTAGTG 406
QY 1105 CATGGGCATCTGACACTGTGTACGTGTCTTTTGGAGCTGGTCCCTCTCTCTGTGTAGC 1164
Db 407 AACGGGCACTTGGAGACACTGAGATCTCTTTTGGAGCTGGTCTGATCCCAACGGCAGC 466
QY 1165 ATCTAACAACACTGTTCTCCCGTCTCACAGCTGCCCGTGTATGGTGTCTGTGTATCCTG 1224
Db 467 CGGCACACCGGAGCACTCTGTGTACATGCTCTCTGTGGTAGGAGCAGCATCCTG 526
QY 1225 CAGGAGCTCTAGACCATGTGTGAGGCGCAACGTCAGAACTAACT-----ACC 1274
Db 527 AAGGCTCTTATCAGGTATGGGCGAGATGTTGATGTCAACCATCATCTGAAATCTGCACAC 586
QY 1275 AGTCTGGGCATCAACATAGTCTCATGTTCT-----GGCCCCCTCTATTGGCGCA 1326
Db 587 CGGCCCTCTTTTTCAGCGGCGCTTAACTCTCTGGTGTCTGTCTCTATATCATCATGCT 646
QY 1327 GTCTACGGGCACTGTGACTGTTTCCGCTCTCTTTCCTCCAGCGGCGAGACCCCTGACTAC 1386
Db 647 GCCTACCATCACTTTCAGTCTCTCAGGCTGCTCTTGGAGCTGGGCGAATCTCTGACTTC 706
QY 1387 AACTGCACTGACGAGGC 1404
Db 707 AATTGCAATGGCCCTGTC 724

RESULT 2

US-09-302-769-26
; Sequence 26, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Human
US-09-302-769-26

Query Match 3.8%; Score 71.2; DB 4; Length 419;
Best Local Similarity 55.4%; Pred. No. 2.2e-11;
Matches 164; Conservative 0; Mismatches 123; Indels 9; Gaps 1;
QY 862 GAGAAGCAGGCTCTCAATCAAGCAGTGTATGACAAGCTCTCTATCTTTGGACCAGCTT 921
Db 114 GAGGACACGAGGCTCCATGATGAGCTTACGTCGGGACCTCCAGACCTCAGGACCTA 173
QY 922 TTGCGCCAGGAGCGTTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGG-----GGT 972
Db 174 TTGCAAGAGGAGAGCTACCGGACCGCATCAACGAGAAGTCTCTCTGGTGTCTGGCTGG 233
QY 973 GTTCTGGGACACCTTGGCTTGGCTTCTTATGGCCACTTGAGCTGTGTCAGATC 1032
Db 234 CTCCTCTGCACACCGTTTGGAAATCGCGGCCACTGCGAGGCAATGGGAGCTGTGTGACTTC 293
QY 1033 CTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGACGCGCACTTTTC 1092

Db 294 CTATCCGGAAGGGGGCCGAGGTGGATCTGTGGACGTAAAGAGCAGACGCGCTGTAT 353
QY 1093 ACTGCTGTCACTATGCGCATGCGCATCTGACTGTGTACGTGTCTTTTGAAGCTGTGC 1148
Db 354 GTGGCTGTGTGAACGGGCACCTAGAGAGTACCAGATCCTTCTCTGAAGCTGCGGC 409
RESULT 3
US-09-221-298-28
; Sequence 28, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (212)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-28

Query Match 3.7%; Score 69.4; DB 3; Length 401;
Best Local Similarity 52.4%; Pred. No. 7.4e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGGAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTAT 419
Db 52 AGAAAAGTTTAACTGGTGGATGGACAGCTTTTACAATCCAGTGAAGAAATACAGGAGGCA 111
QY 420 GGAATACCCAAACAGAACTGGACCATACAGATGCCCAACAGAACTATGAGATATGCAG 479
Db 112 GGGCTTGGCCCAATCACCATTGGAGAACTAACTTTTATTAAGTGTCTATGAGCTCTGCGA 171
QY 480 CACCTACCTCTGAAATAGTGTCTCTAAATCTGTCTACCTTTGGGAACGCTGTGGAAG 539
Db 172 CACTTACCTGTCTCTTTGGTGTCTCTGCTGTATCGTGCCTCANATGATGACTCCGGAGAGT 231
QY 540 TTCAAAGTTTCAAGAGTAAAGACGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Db 232 TCAACTTTTGTAGTCCCGAAATCGAATTCAGTGTCTCATGATTCATCCAGAAATAA 291
QY 600 TGCTGCATTTGCGGTGTAGCCAGCTCTCTCTGGAAT---TTACACTGCTGTGTAGA 656
Db 292 GACGGTCAATTGTGCGTTCAGTCAAGTCCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 351
QY 657 TGATGAGCTCTTGTGGAGGCCATTAGCCAAACAA 692
Db 352 TGATGAGAAATATCTCGATGTTATCAGGAGACTAA 387

RESULT 4

US-09-155-078-1
; Sequence 1, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra
; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia Louis
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078

;/ CURRENT FILING DATE: 1999-03-02
;/ EARLIER APPLICATION NUMBER: PCT/DE97/00592
;/ EARLIER FILING DATE: 1997-03-21
;/ EARLIER APPLICATION NUMBER: DE 196 11 234.6
;/ EARLIER FILING DATE: 1996-03-21
;/ NUMBER OF SEQ ID NOS: 4
;/ SOFTWARE: Patent in Ver. 2.0
;/ SEQ ID NO 1
;/ LENGTH: 3431
;/ TYPE: DNA
;/ ORGANISM: German-type microorganism & cell cul.
;/ FEATURE:
;/ NAME/KEY: CDS
;/ LOCATION: (1)..(1863)
US-09-155-078-1

Query Match 3.6%; Score 67.2; DB 4; Length 3431;
Best Local Similarity 52.1%; Pred. No. 1.3e-09;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGGAAGTGGATGGAAACTGATTGACCCCAATATACAGACTTTGGGGGTAT 419
DB 522 AGAAAAGTTTACGTGGATGGATGGACAGTTTACATCCAGTGGAGAAATACAGAGGCA 581
QY 420 GGAATACCCACAGAAACTGACCAATACAGATGCCAACAGAAACTATGAGATATGCAG 479
DB 582 GGGCTTGCCCAATCACCATTGGAGAAATACTTTTATTATAAGTCTATGAGCTCTGTGA 641
QY 480 CACCTACCTCTGAAATAGTGGTTCCTAAATCTGTACTCTGGGAACGGTGGTGGAG 539
DB 642 CACTTACCTCTCTTTTGGTGGTCCGTATCGTGCCTCAGATGATGACCTCCGGAGAGT 701
QY 540 TTCAAAGTTTCAAGTAAGAAAGCTGCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 599
DB 702 TCGAACTTTAGTCCCGAAATCGAATTCAGAGCTGTGATGATTCATCCAGAAATAA 761
QY 600 TGCTGCCATTTGCCGTGTAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
DB 762 GACGGTCAITGTGCGTGTGAGTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 821
QY 657 TGATGAGCTCTTGTGGAGGCCATTAGCCAAACAA 692
DB 822 TGATGAGAAATATCTGATGTTATCATCAGGAGACTAA 857

RESULT 5
US-08-181-629A-2
; Sequence 2, Application US/08181629A
; Patent No. 5472872
; GENERAL INFORMATION:
; APPLICANT: Swaminathan, Neela
; APPLICANT: Van Etten, James
; APPLICANT: Mead, David
; APPLICANT: Skowron, Piotr
; TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,629A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

;/ NAME: Clough, David W.
;/ REGISTRATION NUMBER: 36,107
;/ REFERENCE/DOCKET NUMBER: 31504
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 312/474-6300
;/ TELEFAX: 312/474-0448
;/ TELEX: 25-3856
;/ INFORMATION FOR SEQ ID NO: 2:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 5496 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
US-08-181-629A-2

Query Match 3.5%; Score 66.6; DB 1; Length 5496;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 144; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 978 TGGGACACCCCTTGGCGCTTGGCTTCTTATGCGCACCTTGAGCTGTTTGCAGTCTCTT 1037
DB 3468 TTGGACGGGTTTACATTTAGCGGCTTTTAAATGTCATAGCATGTGGCTCAAGACGCTTAT 3527
QY 1038 AGCCCATGCTGTGATGTTGACAGCTTGATGTCAAGGCACAGACGCCACTTTTCACTGC 1097
DB 3528 TGATCGGGTGCAATCTTGACATCACAGATATTCGGGATGTACACCACTTCATCGTGC 3587
QY 1098 TGTGAGTCATGCCCATCTGGACTGTGAGTGTGCTGCTTTTGGAGCTGGTCTCTCTGG 1157
DB 3588 GGTTTATAATGACCAAGCATGTCGTAAGATACCTCGTAGAAGCAGGTCGCAACTCTTGA 3647
QY 1158 TGGTAGCATCTACAACTGTTCTCCGCTGCTCACAGCTGCCGTGATGCTGCTTGC 1217
DB 3648 CTTCAATGATGATGCTGAGTGGTGGCTTCACTTACCGGGCTTTTAAATGGTAATGATGC 3707
QY 1218 TATCTCGAGGAGCTCTTAGACCATGTCAGCA 1250
DB 3708 GATTTGAGGATGCTCATTTGAAGCAGGTGCAGA 3740

RESULT 6
US-09-509-802-1
; Sequence 1, Application US/09509802
; Patent No. 6489130
; GENERAL INFORMATION:
; APPLICANT: Immunex Corp.
; APPLICANT: Bird, Timothy
; APPLICANT: Virca, G.D.
; TITLE OF INVENTION: DEATH ASSOCIATED KINASE CONTAINING ANKYRIN REPEATS (DAKAR)
; FILE REFERENCE: 2889-US
; CURRENT APPLICATION NUMBER: US/09/509,802
; CURRENT FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-509-802-1

Query Match 3.2%; Score 60.2; DB 4; Length 2370;
Best Local Similarity 51.3%; Pred. No. 1.2e-07;
Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
QY 980 GGACACCCCTTGGCTTGGCTTCTTATGCGCACCTTGAGCTGTTTGCAGTCTCTTAG 1039
DB 1829 GGACACCCCTTGCACCTGCTGTCTAGAGGGGCAATTACCGTGGCTCGCATCTCTATTG 1888
QY 1040 CCCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGCTG 1099
DB 1889 ACCTGTGCTCTGATGTTAAATCTGCGCCTACAGGCACAGACACCTCTGTCATGTTGCTG 1948

Db 3875 CCAAAGCTGGCCTTGGAGGTGGTCAAGTTGCTGCGAGGCGGTGCTCGCCAAAT 3934
QY 1160 GTAGCATCTACACAACTGTTCTCCGCTGCTACAGCTCCCGTGATGCTGTGCTA 1219
Db 3935 CGAGACCAACTACGGTTGCGCGCCATTGTTGTCGCCGCTCCGAGGACACAACGAGG 3994
QY 1220 TCTGCGAGAGTCTCTAGACCATGGTGCAGAGGCCAACGTC 1260
Db 3995 TCTGCGGTATCTGATGACAGGACGACACCTACGGC 4035

RESULT 10

US-09-392-812A-1
; Sequence 1, Application US/09392812A
; Patent No. 6537778
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Walker, Richard G.
; APPLICANT: Willingham, Aaron
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: A Eukaryotic Mechanosensory Transduction Channel
; FILE REFERENCE: 02307E-097600US
; CURRENT APPLICATION NUMBER: US/09/392,812A
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 24358
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: genomic nompC (no-mechanoreceptor potential C)
; OTHER INFORMATION: nucleotide sequence
US-09-392-812A-1

Query Match 2.9%; Score 55.4; DB 4; Length 24358;

Best Local Similarity 49.8%; Pred. No. 1.4e-05;

Matches 140; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 980 GGACACCTTGGCTTGGCTGCTTATGCGCACTTGAGTGTGTCGAAGTCTCTTAG 1039
Db 13889 GGACGGGCTGCATATCGCGCCATGATGATGCGCCACATCCAGATGTTGGAGATTCTGCTG 13948
QY 1040 CCATGCTGCTGATGTTGACAGCTTGATGCTCAAGCAGACGACGACCTTTCACTGCTG 1099
Db 13949 GCGAGGCGGAGATCAACGCAACCGATCGGAACGGTTGGACGCACTGCAATTGTGCTG 14008
QY 1100 TCAGTCATGCGCATCTGGAAGTGTGCTGCTGCTTTTGAAGTGTGCTCTCTCTGCTG 1159
Db 14009 CCAAAGCTGGCCTTGGAGGTGCTGAGTTGCTGCGAGGCGGTGCTCGCCAAAT 14068
QY 1160 GTAGCATCTACACAACTGTTCTCCGCTGCTACAGCTGCCGCTGATGCTGTGCTA 1219
Db 14069 CGAGACCAACTACGGTTGCGCGCCATTGTTGTCGCCGCTCCGAGGACACAACGAGG 14128
QY 1220 TCTGCGAGAGTCTTAGACCATGGTGCAGAGGCCAACGTC 1260
Db 14129 TCTGCGGTATCTGATGAACAGGACGACACCTACGGC 14169

RESULT 11

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-Fls
US-08-232-463-14

Query Match 2.8%; Score 51.8; DB 1; Length 7218;

Best Local Similarity 7.4%; Pred. No. 8e-05;

Matches 32; Conservative 218; Mismatches 185; Indels 0; Gaps 0;

QY 1247 CAGAGCCCAACGCTCAAAAGTAACTACCACTCTGGGCATCAAAACATAGCTTCTG 1306
Db 1020 CACAGAAATTAATTCGAGCTTGCTGCGAGTTCGAGGAGCTTCGATATATATATAT 1079
QY 1307 GCGCCCTCTATTGGCGCAGTCTACGGGCACCTGAGACTGTTCCGCTGCTTTGCTCC 1366
Db 1080 YY 1139
QY 1367 ACGGGCAGACCTGACTACAACTGACCTGACGAGGCGCTATTGGCTGCTGTCGAAGC 1426
Db 1140 YY 1199
QY 1427 CCGCACCCTCTGAAATCTGCTCCATCAATAATTGTGAGCCAGAGTATATCCAGCTGT 1486
Db 1200 YY 1259
QY 1487 TAATCGATTTGGTGTGCTAATATCTACCTTCAATCTCTCCCTTGACCTGACCTCAAG 1546
Db 1260 YY 1319
QY 1547 ATGATAAAGGCATTGCATTGCTGCTACAGGCCGAGCCACTCCACGGTCACTTCTATCAC 1606
Db 1320 YY 1379
QY 1607 AGTCCGTTTAGTCCGCGAGAGCTTGTGCGAGGCTGCCAGCCAGCCAGCATCAACC 1666
Db 1380 YY 1439
QY 1667 AGCTGGATATCTCTC 1681
Db 1440 AAATCTCTCTATCTC 1454

RESULT 12

US-09-196-387-1
; Sequence 1, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; US-09-196-387-1

Query Match 2.7%; Score 51.4; DB 3; Length 4134;
Best Local Similarity 49.1%; Pred. No. 7.5e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCCTTGGCTGGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCTCTTAGCCCC 1042
DB 658 CTCCTCGCACTTCGCTGCAGGTTTGGGAAGGAAGGATGTTGTAGAACACTTACTACAGA 717
QY 1043 ATGCTGCTGATGTGACAGCTTGAGTGTCAAGCAGAGCCACTTTTTCACCTGCTGTCA 1102
DB 718 TGGGTGCTAATGTCCACGCTCGTGATGATGAGGCTCTCATCCCGCTTCAATAGCCTGTT 777
QY 1103 GTCATGGCCATCTGGACTGTGTACGTGTGCTTTTGGAAAGCTGCTCTCTCTGGTGGTA 1162
DB 778 CTTTGGCCATGCTGAGTTGTGAGTCTGTTATGTGCCAAGGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACAACTGTTCTCCCGTGTCTACAGCTGCCCGTGTGCTGTGCTGTATCC 1222
DB 838 GGGATAACTGGAACATAACACCTCTGCATGAAGCTGCTATTAAAGGAAGATCGATGTGT 897
QY 1223 TGCAGAGCTCTAGACCATGTTGTCAGAGGCCAAAGT 1259
DB 898 GCATTGTGCTGTCAGCAGGAGCTGACCCAAACAT 934

RESULT 13

US-09-841-835-1
; Sequence 1, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; US-09-841-835-1

Query Match 2.7%; Score 51.4; DB 4; Length 4134;
Best Local Similarity 49.1%; Pred. No. 7.5e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCCTTGGCTGGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCTCTTAGCCCC 1042
DB 658 CTCCTCGCACTTCGCTGCAGGTTTGGGAAGGAAGGATGTTGTAGAACACTTACTACAGA 717
QY 1043 ATGCTGCTGATGTGACAGCTTGAGTGTCAAGCAGAGCCACTTTTTCACCTGCTGTCA 1102
DB 718 TGGGTGCTAATGTCCACGCTCGTGATGATGAGGCTCTCATCCCGCTTCAATAGCCTGTT 777
QY 1103 GTCATGGCCATCTGGACTGTGTACGTGTGCTTTTGGAAAGCTGCTCTCTCTGGTGGTA 1162
DB 778 CTTTGGCCATGCTGAGTTGTGAGTCTGTTATGTGCCAAGGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACAACTGTTCTCCCGTGTCTACAGCTGCCCGTGTGCTGTGCTGTATCC 1222
DB 838 GGGATAACTGGAACATAACACCTCTGCATGAAGCTGCTATTAAAGGAAGATCGATGTGT 897
QY 1223 TGCAGAGCTCTAGACCATGTTGTCAGAGGCCAAAGT 1259
DB 898 GCATTGTGCTGTCAGCAGGAGCTGACCCAAACAT 934

RESULT 14

US-09-196-387-7

; Sequence 7, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2027
; US-09-196-387-7

Query Match 2.7%; Score 51.4; DB 3; Length 4491;
Best Local Similarity 49.1%; Pred. No. 7.9e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCCTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCCTCTTAGCCC 1042
Db 658 CTCCTCGCACTTCGCTGAGGTTTGGGAAGGAGGATGTTGAGAACACTTACTACAGA 717
QY 1043 ATGGTGCTGATGTGACAGTTGGATGTCAAGGCACAGAGCCACTTTTCACTGTGTCA 1102
Db 718 TGGGTGCTAATGTCACCGCTCGTGATGGAGGCTCATCCCGCTTCATAATGCTGTT 777
QY 1103 GTCATGGCCATCTGAGCTGTGTACGTGTGCTTTTGGAGAGCTGGTGTCTCTCTGGTGTGA 1162
Db 778 CTTTGGCCATGCTGAGTTGTGAGTCTGTTATTTGGCCAGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACTGTTCTCCCGTGCTCACAGCTGCCGCTGATGTGTGCTGTGTATCC 1222
Db 838 GGGATAACTGGAACTATACACCTCTGCATGAAGCTGCTATTAAAGGGAAGATCGATGCT 897
QY 1223 TGCAGAGCTCCTAGACCATGTGTGAGAGGCCCAAGCT 1259
Db 898 GCATTGTGCTGTGCAGCAGGAGCTGAGCCCAACAT 934

RESULT 15

US-09-841-835-7
; Sequence 7, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2027
; US-09-841-835-7

Query Match 2.7%; Score 51.4; DB 4; Length 4491;
Best Local Similarity 49.1%; Pred. No. 7.9e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCCTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCCTCTTAGCCC 1042
Db 658 CTCCTCGCACTTCGCTGAGGTTTGGGAAGGAGGATGTTGAGAACACTTACTACAGA 717
QY 1043 ATGGTGCTGATGTGACAGCTTGGATGTCAAGGCACAGAGCCACTTTTCACTGTGTCA 1102
Db 718 TGGGTGCTAATGTCACCGCTCGTGATGGAGGCTCATCCCGCTTCATAATGCTGTT 777
QY 1103 GTCATGGCCATCTGAGCTGTGTACGTGTGCTTTTGGAGAGCTGGTGTCTCTCTGGTGTGA 1162
Db 778 CTTTGGCCATGCTGAGTTGTGAGTCTGTTATTTGGCCAGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACTGTTCTCCCGTGCTCACAGCTGCCGCTGATGTGTGCTGTGTATCC 1222
Db 838 GGGATAACTGGAACTATACACCTCTGCATGAAGCTGCTATTAAAGGGAAGATCGATGCT 897
QY 1223 TGCAGAGCTCCTAGACCATGTGTGAGAGGCCCAAGCT 1259

Db 898 GCATTGTGCTGCTGCAGCAGGAGCTGACCCAAACAT 934

Search completed: December 1, 2003, 13:33:36
Job time : 141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 2, 2003, 08:30:46 ; Search time 0.001 Seconds
(without alignments)
1686.444 Million cell updates/sec

Title: us-09-941-831a-20
Perfect score: 2343
Sequence: 1 MRESGWKLDIPIDFGRMG1.....QAINQLDIPPLISYLKHL 449

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1 seqs, 1878 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=soft -Q=us-09-941-831a-20 -DB=us-09-941-831a-6
-SUFFIX=pro -OUT=align20_6 -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits
-START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1 -DOCALIGN=200
-THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO XLPXY
-NEG SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXIT=7

Database : us-09-941-831a-6:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length DB	ID	Description
1	2343	100.0	1878	1 us-09-941-831a-6

ALIGNMENTS

RESULT 1
us-09-941-831a-6

Alignment Scores:

Pred. No.:	0	Length:	1878
Score:	2343.00	Matches:	449
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

us-09-941-831a-20 (1-449) x us-09-941-831a-6 (1-1878)

Qy	1	MetArgLusertGlyTrpLysLeuIleAspProIleSerAspPheGlyArgMetGlyIle	20
Db	367	ATGAGGGAAGTGGTGGAACTGATTGACCAATATCAGACTTTGGGCGTATGGGAATA	426

Qy	21	ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr	40
Db	427	CCCAACAGAACTGGACCATACACATGCCAACAGAACTATGAGATATGCAGACCTAC	486
Qy	41	ProProGluIleValProLysSerValThrLeuGlyThrValValGlySerSerLys	60
Db	487	CCTCTGAAATAGTGGTTCCTAAATCTGTACCTTGGGAACGGTGGTTGGAAGTTCAAA	546
Qy	61	PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla	80
Db	547	TTCAAGAGTAAAGAACGGTCTCCTGTCTCTCACTCTCAAAAGAGAACAAATGCTGCC	606
Qy	81	IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu	100
Db	607	ATTTCGGCTGTAGCCAGCCTCTCTCTGGAATTTTACACTCGCTGTGTAGATGATGAGCTC	666
Qy	101	LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr	120
Db	667	TTGTTGGAGGCCATTTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA	726
Qy	121	ArgProLysIleTrpPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys	140
Db	727	AGACCAAGATCTGGCATTTCTGTGTCTCAATAGAAATAGTCTTCCAAATTAGCCAAG	786
Qy	141	MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu	160
Db	787	ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCGCAGCCCGACAGAGGAGGAG	846
Qy	161	AspThrAspThrGluGluGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr	180
Db	847	GACACTGACACAGAGGAGGAGCAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTAT	906
Qy	181	ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly	200
Db	907	ACTTTGGACCACTTTTGGCCAGGAGCGTTTCAAAACGTTTCAACACGAGGAGTGGC	966
Qy	201	TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeu	220
Db	967	TGGGTGTCTCTGGGACACCTTGGCTGGCTTCTTATGGCCACTTGGAGCTGTTTG	1026
Qy	221	GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro	240
Db	1027	CAAGTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGACGCCA	1086
Qy	241	LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly	260
Db	1087	CTTTTCACTGTCTGTCAGTCATGGCCATCTGGACTGTGTACGTGTCTTTTGGAAAGCTGT	1146
Qy	261	AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp	280
Db	1147	GCCTCTCTCTGGTGGTAGCATCTACAAACACTGTTCTCCGCTCTCACAGCTCCCGTAT	1206
Qy	281	GlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAla	300
Db	1207	GGTGTGTGTCTATCTTCAGGAGCTCTTAGACCATGGTGCAGAGGCCCAACGCTCAAAGCT	1266
Qy	301	LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla	320
Db	1267	AAACTACCAGTCTGGGCATCAAAACATAGCTTCTGTGGCCCTCTTATTTGGCCGCA	1326
Qy	321	ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyValAspProAspTyr	340
Db	1327	GTCACGGGCACCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGCAGACCTGACTAC	1386
Qy	341	AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle	360
Db	1387	AACTGCACTGACAGGGCTATTGGCTGGTGTCCCAAGACCCCGCACCTCTCTTGAATC	1446
Qy	361	CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn	380
Db	1447	TGCCTCCATCAATAATGTGAGCCAGATATCCAGCTGTAAATCGATTTTGTGTGTAAT	1506

```
Qy 381 ILeTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
Db 1507 ATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAAGATGATAAAGGCATTGCATTG 1566
Qy 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
Db 1567 CTGCTACAGGCCCGAGCCACTCCACGGTCACTTCTATCACAGGTCCGTTTAGTCGTCGCGC 1626
Qy 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProPromet 440
Db 1627 AGAGCCTTGTGCCAGGCTGCCGCCACAGCCATCAACCAGCTGGATATTCTCTCCCATG 1686
Qy 441 LeuIleSerTyrLeuLysHisGlnLeu 449
Db 1687 TTGATTAGCTACCTATAAACCACTG 1713
```

Search completed: December 2, 2003, 08:30:48
Job time : 2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2003, 09:49:33 ; Search time 3611 Seconds

(without alignments)

12640.212 Million cell updates/sec

Title: US-09-941-831A-6

Perfect score: 1878

Sequence: 1 catgattacccaagcttgg.....taaaaaaaaaaaaaaaaaa 1878

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estnu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_man:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	729.2	38.8	1262	11 AK009644	AK009644 Mus muscu
2	675.2	36.0	1201	9 AL529411	AL529411 AL529411
3	644	34.3	644	13 BX107270	BX107270 BX107270
4	617	32.9	625	10 AW957331	AW957331 EST369521

5	593.8	31.6	719	12	BM190670
6	532.2	28.3	943	10	BF037514
7	498.8	26.6	678	29	AG127205
8	482	25.7	482	9	AA633985
9	452.2	24.1	543	14	W72063
10	445	23.7	446	14	W76256
11	404.6	21.5	464	14	W58499
12	388.8	20.7	579	12	BI344684
13	368.6	19.6	859	9	AJ454588
14	368.4	19.6	518	28	AZ725513
15	358	19.1	362	9	AA349955
16	344.2	18.3	932	14	BY709303
17	339	18.1	342	9	AA349954
18	332.2	17.7	751	13	BU383527
19	331	17.6	643	10	BB661935
20	330.2	17.6	722	9	AJ451253
21	329	17.5	627	10	BB662275
22	314.8	16.8	659	12	BM439955
23	310.8	16.5	908	13	BU220575
24	297.6	15.8	524	14	W58534
25	297.4	15.8	861	13	BU254916
26	293.8	15.6	538	12	BI290522
27	279.4	14.9	372	28	AZ725227
28	276	14.7	849	13	BU228932
29	275	14.6	713	14	CA364256
30	269.8	14.4	486	10	BF889720
31	269.4	14.3	758	13	BX071379
32	265.8	14.2	755	14	CA368575
33	265.4	14.1	680	14	CA358977
34	258.8	13.8	990	12	BM545864
35	249.2	13.3	318	9	AA778891
36	241.4	12.9	1060	29	CC217390
37	233.6	12.4	2165	11	AK081973
38	232.6	12.4	798	13	BQ179015
39	229.6	12.2	567	10	BF949440
40	228	12.1	3814	11	AK076218
41	226.4	12.1	837	13	BQ946838
42	226	12.0	690	13	BU382590
43	224.2	11.9	2471	11	AK028160
44	223.8	11.9	436	12	BI290849
45	223.8	11.9	464	10	BF555515

ALIGNMENTS

RESULT 1
AK009644
LOCUS
DEFINITION
Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310036C05 product:ankyrin repeat domain-containing SOCS box protein 12, full insert sequence.

AK009644 1262 bp mRNA linear HTC 05-DEC-2002

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

AK009644.1 GI:12844562

Db 625 TTGCCATCTTACAGGAGCTCTAGGCGATGGTCCGAGGCTAATGTCAAAGCTAAACTAC 684
Qy 1274 CAGTCTGGGATCAAAATAGCTTCATGTTCTGGCCCTCTATTTGGCGGAGTCTACG 1333
Db 685 CAGTCTGGGCGTCAAAATAGCTTCATGTTCTGGCCCTCTATTTGGCGGAGTCTATG 744
Qy 1334 GGCACCTGGAAGTCTTTCCGCTCTTTTGTCTCCAGCGGGCAGACCTGACTACAACTGCA 1393
Db 745 GGCACCTTGATGTTTCCGCTCTTTTGTCTATGCGGCGAGATCTGATTACAACTGCA 804
Qy 1394 CTGACAGGGCCTATTGGCTGCTGTCCTCCAAAGACCCGCGACCTCTCTGAAATCTGCTCC 1453
Db 805 CTGACAGGGCCTTTTAAAGTCTGTTCTCCAGCGCTCGCACCTCTCTGAAATCTGCTTC 864
Qy 1454 ATCATATTTGTAGCCAGAGTATATCCAGCTGTATATCGATTTTGGTGTATATCTACC 1513
Db 865 ATCATATTTGTAGCCAGAGTATATCCAGCTGTATATCGATTTTGGTGTATATCTACC 924
Qy 1514 TTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAATGCAATGCTCTAC 1573
Db 925 TTCCATCTCTCTCTGTCGACCTCAAGATGATAAGGCAATGCAATGCTCTAC 984
Qy 1574 AGCCCGAGCACTCCAGGCTCACTCTATCAAGTTCGTTTGTGCTCCGAGACCT 1633
Db 985 AGCCCGAGCACTCCAGGCTCACTCTCTGTCGACCTCAAGATGATAAGGCAATGCTCTAC 1044
Qy 1634 TGTGCGAGGCTGCGCCAGCCACAGGCTCACTCTATCAAGTTCGTTTGTGCTCCGAGACCT 1693
Db 1045 TGTGCGAGGCTGCGCCAGCCACAGGCTCACTCTATCAAGTTCGTTTGTGCTCCGAGACCT 1104
Qy 1694 GCTACCTAAACCACTGATCTGATCTGAGCTGCTCCAGGACTT-----ATGATGCTT 1748
Db 1105 GCTACCTAAACCACTGATCTGATCTGAGCTGCTCCAGGACTT-----ATGATGCTT 1164
Qy 1749 CGAAAA--CCACTGGGAGCTACGAGTGTGGAGAGCATTTACAGCCTCATCCACTTACCT 1807
Db 1165 CAAAAATAAACCTGTGATCCAGGAGTACTGTAGTGAATAATACACCTCATACATTTCT 1224
Qy 1808 GGAG 1811
Db 1225 GGCG 1228

RESULT 2
AL529411 1201 bp mRNA linear EST 23-MAY-2003
LOCUS AL529411 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DD006Y19 5-PRIME, mRNA sequence.
ACCESSION AL529411
VERSION AL529411.2 GI:31067254
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1201)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12792904.
Contact: Genoscope
Genoscope, Centre National de Sequenage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10696. r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DD006Y19&cluster=10696. r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DD006Y19Q1.

FEATURES
source Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DD006Y19"
/issue_type="NEUROBLASTOMA COT 50-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 309 a 241 c 276 g 326 t 49 others
ORIGIN
Query Match 36.0%; Score 675.2; DB 9; Length 1201;
Best Local Similarity 95.3%; Pred. No. 1.2e-96;
Matches 682; Conservative 20; Mismatches 13; Indels 1; Gaps 1;
Qy 23 CGAGGTAGAAAACGTGAAATTTGGTGGATCGTTATGTAGTAAGAAACAGCTAATGGGA 82
Db 136 CCAGGTAGAAAACGTGAAATTTGGTGGATCGTTATGTAGTAAGAAACAGCTAATGGGA 195
Qy 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGCTTTCAGGTGCGAGCCCGAAG 142
Db 196 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGCTTTCAGGTGCGAGCCCGAAG 255
Qy 143 AAACATGATGTCACCTCCATCATATGCCATGTGGAGAGTTACCCATCAGTACGCTGG 202
Db 256 WWACATGATGTCACCTCCATCATATGCCATGTGGAGAGTTACCCATCAGTACGCTGG 315
Qy 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTTCCGGTGGCCCACTTTGTTTGTAGATTCTG 262
Db 316 GTTGTCCCTGACCTCCGCTGCAAGAAATTTCCGGTGGCCCACTTTGTTTGTAGATTCTG 375
Qy 263 ACCTTGTGTGCAATGAGGTTTATTTTCACTGCTCAAGCTTTTCTCAGCGAGCATTTACCTG 322
Db 376 ACCTTGTGTGCAATGAGGTTTATTTTCACTGCTCAAGCTTTTCTCAGCGAGCATTTACCTG 435
Qy 323 AAGATCTTTATGCTTTTCTTATATCCCAATCTCCAAAGAGATGAGGGAAGTGGAT 382
Db 436 AAGATCTTTATGCTTTTCTTATATCCCAATCTCCAAAGAGATGAGGGAAGTGGAT 495
Qy 383 GCAAACTGATGACCCCAATATCAGACTTTGGGCGTATGGGAATACCCACAGAACTGGA 442
Db 496 GGTAACTGATGACCCCAATATCAGACTTTGGGCGTATGGGAATACCCACAGAACTGGA 555
Qy 443 CCATAACAGATGCCAACAGAAATATGAGATATGAGATATGCTCTCTGAAATAGTG- 501
Db 556 CCATAACAGATGCCAACAGAAATATGAGATATGAGATATGCTCTCTGAAATAGTG- 615
Qy 502 GTTCTAAATCTGTTACCTTGGGAAACGGTGGTGGAAAGTTCAAGTTCAAGAGTAAAGAA 561
Db 616 GTTCTAAATCTGTTACCTTGGGAAACGGTGGTGGAAAGTTCAAGTTCAAGAGTAAAGAA 675
Qy 562 CGTGTCCCTGCTCTCTCTACCTCTACAAAGAGAAATGCTGCCATTTGCCGCTGTACG 621
Db 676 CGTGTCCCTGCTCTCTCTACCTCTACAAAGAGAAATGCTGCCATTTGCCGCTGTACG 735
Qy 622 CAGCTCTCTCTGGAATTTTACACTCGCTGTGATGATGAGCTCTTGTGGAGGCAAT 681
Db 736 CAGCTCTCTCTGGAATTTTACACTCGCTGTGATGATGAGCTCTTGTGGAGGCAAT 795
Qy 682 AGCCAAACAAACCCAGGAGGAGGAGTTTATGTTGTGTAGACACAGACCAAGAT 737
Db 796 AGCCAAACAAACCCAGGAGGAGGAGTTTATGTTGTGTAGACACAGACCAAGAT 851
RESULT 3
BX107270 644 bp mRNA linear EST 06-FEB-2003
LOCUS BX107270 Soares fetal heart NbhH19W Homo sapiens cDNA clone
DEFINITION IMAGP998N09778 ; IMAGE:341192, mRNA sequence.


```

QY 78 TGGGATCTTTATCTTACTGCAACCACTGATCTATGTGGAGGCTTCAGGTGCGAGCCG 137
Db 61 TGGGATCTTTATCTTACTGCAACCACTGATCTATGTGGAGGCTTCAGGTGCGAGCCG 120
QY 138 GAAAGAAACATGGATGCACTGCATCACAATGCGCACTGTGGAGATTAACCATCACTAG 197
Db 121 GAAAGAAACATGGATGCACTGCATCACAATGCGCACTGTGGAGATTAACCATCACTAG 180
QY 198 CTGGGTTGTCCTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTGGCCCACTTTGTTTGA 257
Db 181 CTGGGTTGTCCTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTGGCCCACTTTGTTTGA 240
QY 258 TTCTGACCTTGTGTGCATGAGGTTTATATTTTCACTGCTCAAGCTTTCTCAGCCAGCAT 317
Db 241 TTCTGACCTTGTGTGCATGAGGTTTATATTTTCACTGCTCAAGCTTTCTCAGCCAGCAT 300
QY 318 ACCTGAAGATCTTTATGCTTTTCTTATATATCCAAATCCTCAAAAGAGATGAGGGAAG 377
Db 301 ACCTGAAGATCTTTATGCTTTTCTTATATATCCAAATCCTCAAAAGAGATGAGGGAAG 360
QY 378 TGGATGGAACATGATTGACCCCAATATCAGACTTTGGGCGTGTGGGAATACCCAAAGAAA 437
Db 361 TGGATGGAACATGATTGACCCCAATATCAGACTTTGGGCGTGTGGGAATACCCAAAGAAA 420
QY 438 CTGGACCATTAACAGATGCCAACAGAACTATGAGATATGAGCACCCTACCCCTCTGAAAT 497
Db 421 CTGGACCATTAACAGATGCCAACAGAACTATGAGATATGAGCACCCTACCCCTCTGAAAT 480
QY 498 AGTGGTTCTTAATCTGTTACCTTGGGAACGGTGGTGGAGTTCAAGTTTGAAGTAA 557
Db 481 AGTGGTTCTTAATCTGTTACCTTGGGAACGGTGGTGGAGTTCAAGTTTGAAGTAA 540
QY 558 AGAACCTGTCCCTGTCTCTCTACCTCTACAAAGAGAACTGCTGCCATTTGCCGCTG 617
Db 541 AGAACCTGTCCCTGTCTCTCTACCTCTACAAAGAGAACTGCTGCCATTTGCCGCTG 600
QY 618 TAGCCAGCCTCTCTGGAATTTTAC 642
Db 601 GAGCCAGCCTTCTGGAATTTTAC 625

```

```

RESULT 5
BM190670 719 bp mRNA linear EST 11-DRC-2001
LOCUS POSM0100011.E08F porcine skeletal muscle cDNA library (POSF) Sus
DEFINITION scrofa cDNA 5', mRNA sequence.
ACCESSION BM190670
VERSION BM190670.1 GI:17526633
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

```

```

REFERENCE 1 (bases 1 to 719)
AUTHORS Yao,J., Coussens,P., Ernst,C. and Saama,P.
TITLE Analysis of expressed sequence tags from a normalized porcine
skeletal muscle cDNA library
JOURNAL Unpublished
COMMENT Contact: Jianbo Yao
Department of Animal Science
Michigan State University
B215 Anthony Hall, East Lansing, MI, USA
Tel: 517-353-8443
Fax: 517-353-1699
Email: yaoj@msu.edu
Seq primer: M13 reverse.

```

```

FEATURES
source
1..719
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/sex="male and female"
/tissue_type="skeletal muscles"

```

```

/dev_stage="45 d and 90 d of gestation, birth, 7 wk and 1
yr of age"
/lab_host="DH10B"
/clone_lib="porcine skeletal muscle cDNA library (POSF)"
/notes="organ: hind limbs; Vector: pSPORT1; Site_1: NotI;
Site_2: SalI"
BASE COUNT 151 a 206 c 176 g 182 t 4 others
ORIGIN
Query Match 31.6%; Score 593.8; DB 12; Length 719;
Best Local Similarity 90.0%; Pred. No. 9.1e-84;
Matches 645; Conservative 0; Mismatches 71; Indels 1; Gaps 1;
QY 818 CCTCTCTGAGCCGACAAAGGAGGAGGAGGACACTGACACAGAGAGAGAGAGGCTCTCA 877
Db 2 CACGGCTCGGCGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 61
QY 878 ATCAAGCAGTGTATGACAAACGACTCCTATATCTTTTGACACGAGCTTTTGCCGAGGCGTT 937
Db 62 ATCAAGCAGTGTATGACAAACGACTCCTATATCTTTTGACACGAGGAGGAGGAGGAGG 121
QY 938 ACAAAGCTTTTCAATCAACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 996
Db 122 ACAAAGCTTTTCAATCAATAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 181
QY 997 GCTGCTTCTTATGGCCACTTGTAGCTGTTTGCAGAGTCTCTTAGCCCATGCTGATGTT 1056
Db 182 GCTGCTTCTTATGGCCACTTGTAGCTGTTTGCAGAGTCTCTTAGCCCATGCTGATGTT 241
QY 1057 GACAGCTTGGATGTCAAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
Db 242 GACAGCTTGGAGGTCAAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
QY 1117 GACTGTGACGTGCTGTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
Db 302 GACTGTGACGTGCTGTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 361
QY 1177 TGTTCTCCCGTCTCACAGCTCCCGTGTATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1236
Db 362 TGCTCTCCCGTCTCACAGCTCCCGTGTATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 421
QY 1237 GACCATGTGTGAGAGGCGCAACGTCAAAGCTAACTACCACTGACAGGAGGAGGAGGAGG 1296
Db 422 GGTATGTGTGAGAGGCGCAATGTCAAGGCAAACTACCACTGCTGGTCAATCAAACTAGCT 481
QY 1297 TCATGTTCTGGGCGGCTCTATTTGGCGGAGTCTACGGGAGGAGGAGGAGGAGGAGG 1356
Db 482 TCATGTTCTGGGCGGCTCTATTTGGCGGAGTCTACGGGAGGAGGAGGAGGAGGAGG 541
QY 1357 CTTTGTCTCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1416
Db 542 CTTTGTCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 601
QY 1417 GTCCCAAGACCCGCGACCCCTCTTGAATCTGCCTCCATCATATAATTTGTGAGGAGGAG 1476
Db 602 GTCCCAAGACCCGCGACCCCTCTTGAATCTGCCTCCATCATATAATTTGTGAGGAGGAG 661
QY 1477 ATCCAGCTGTTAATCGATTTTGGTGAATATATATCTACCTTCCATCTCTCTCCCTGAC 1533
Db 662 ATCCAGCTGTTAATCGATTTTGGTGAATATATATCTACCTTCCATCTCTCTCCCTGAC 718

```

```

RESULT 6
BF037514
LOCUS 601460634F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864128 5',
DEFINITION mRNA sequence.
ACCESSION BF037514
VERSION BF037514.1 GI:10745763
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

QY 910 TTGGACCACTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAGGAGTGCGTGG 969
    |||||
Db 406 TTGGACCACTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAGGAGTGCGTGG 347

QY 970 GGTGTTCTTGGGACACCCCTTGGCTTGGCTGCTTCTTATGGCCACTTGGAGCTGTTTGC 1029
    |||||
Db 346 GGGCTTCTTGGGACACCCCTTGGCTTGGCTGCTTCTTATGAACACTTGGAGCTGTTTGC 287

QY 1030 GTCCCTTTAGCCCATGCTGATGTTGACAGCTTGTCAAGTGTCAAGSCACAGACGCCACTT 1089
    |||||
Db 286 GTCCCTTTAGCCCATGCTGATGTTGACAGCTTGTCAAGTGTCAAGSCACAGACGCCACTT 227

QY 1090 TTCACTGCTGTGAGTCAATGCCATCTGAGCTGTGTGATGCTGTTTGAAGCTGCTGCC 1149
    |||||
Db 226 TTCACTGCTGTGAGTCAATGCCATCTGAGCTGTGTGATGCTGTTTGAAGCTGCTGCC 167

QY 1150 TCTCCTGTTGTTAGTCAATCTACAACTGTTTCTCCGCTGCTACAGCTGCCCGTGATGTT 1209
    |||||
Db 166 TGTCTCTGTTGTTAGTCAATCTACAACTGTTTCTCCGCTGCTACAGCTGCCCGTGATGTT 107

QY 1210 GCTGTTGCTATCTCGCAGGAGCTCTAGACCATGCTGC 1247
    |||||
Db 106 GCTGTTGCTATCTCGCAGGAGCTCTCGCAACGCTCGTAC 69

RESULT 8
AA633985/c
LOCUS
DEFINITION
ac33f04.s1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone
IMAGE:858271 3', mRNA sequence.
ACCESSION
AA633985
VERSION
AA633985.1 GI:2557199
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 482)
AUTHORS
Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
TITLE
Unpublished
JOURNAL
Contact: Wilson RK
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1829 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 446.
FEATURES
source
1. 482
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:858271"
/dev_stage="hnt neurons"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene hnt neuron (#937233)"
/notes="vector: pBluescript SK-; Site 1: EORI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentiated, post mitotic hnt neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
CAATCGGCAGCAG 3' ~3' adaptor sequence: 3'
CTCAGCTTTTCTTTTCTTTT 3'"
BASE COUNT 121 a 88 c 153 g 120 t
ORIGIN

```

```

Query Match 25.7%; Score 482; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 3.9e-66;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1374 AGACCTGCTACAACTGACCTGACCTGACAGGGCTATTGGCTGGTGTCCCAAGACCCCGCAC 1433
    |||||
Db 482 AGACCTGCTACAACTGACCTGACCTGACAGGGCTATTGGCTGGTGTCCCAAGACCCCGCAC 423

QY 1434 CTTCTCTGAAATCTCCCTCCCATCAATAATTGTGAGCCAGAGTATATCCAGCTGTTAATCGA 1493
    |||||
Db 422 CTTCTCTGAAATCTCCCTCCCATCAATAATTGTGAGCCAGAGTATATCCAGCTGTTAATCGA 363

QY 1494 TTTTGGTGTCTAATATCTACCTTTCCATCTCTCTCCCTTGACCTGACCTCACAAGATGATAA 1553
    |||||
Db 362 TTTTGGTGTCTAATATCTACCTTTCCATCTCTCTCCCTTGACCTGACCTCACAAGATGATAA 303

QY 1554 AGGCATTTGATTTGCTGTACAGGCCCGAGCCACTCCAGGTCACTTCTATCAAGTCTCG 1613
    |||||
Db 302 AGGCATTTGATTTGCTGTACAGGCCCGAGCCACTCCAGGTCACTTCTATCAAGTCTCG 243

QY 1614 TTTAGTTCGTCGCGCAGAGCTTGTGCGAGCTGGCCAGCCACCAAGCATCAACAGCTGGA 1673
    |||||
Db 242 TTTAGTTCGTCGCGCAGAGCTTGTGCGAGCTGGCCAGCCACCAAGCATCAACAGCTGGA 183

QY 1674 TATTCCTCCCATGTTGATTAGCTACTAAACACCAACTGTAATCTTGCACTCTCCCGAG 1733
    |||||
Db 182 TATTCCTCCCATGTTGATTAGCTACTAAACACCAACTGTAATCTTGCACTCTCCCGAG 123

QY 1734 GAACCTATGATGCTCCGAAACACCTGGGAGCTCAGTAGCTGGAGAGCATTTACAGCC 1793
    |||||
Db 122 GAACCTATGATGCTCCGAAACACCTGGGAGCTCAGTAGCTGGAGAGCATTTACAGCC 63

QY 1794 TCATCCACTTACCTGGAGCTGCTCTCTGTTATTTATCTCCCAATAAATTTCTCCAGAA 1853
    |||||
Db 62 TCATCCACTTACCTGGAGCTGCTCTCTGTTATTTATCTCCCAATAAATTTCTCCAGAA 3

QY 1854 AT 1855
    ||
Db 2 AT 1

RESULT 9
W72063/c
LOCUS
DEFINITION
zd66f07.s1 Soares fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:345637 3', mRNA sequence.
ACCESSION
W72063
VERSION
W72063.1 GI:1382333
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 543)
AUTHORS
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Woldmann,P. and Wilson,R.
The WashU-Merck EST Project
TITLE
Unpublished
JOURNAL
Contact: Wilson RK
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 945 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 360.
FEATURES
Location/Qualifiers

```

```

source
1. .543
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1271012"
/db_xref="taxon:9606"
/clone="IMAGE:345637"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBHL19W"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCGAAGTGGAGCGCGGCATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Patima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
BASE COUNT 142 a 106 c 164 g 127 t 4 others
ORIGIN
Query Match 24.1%; Score 452.2; DB 14; Length 543;
Best Local Similarity 94.3%; Pred. No. 1.8e-61;
Matches 510; Conservative 0; Mismatches 27; Indels 4; Gaps 4;
QY 1317 TTTGGCGGCGAGTCTACGGGCACCTGGAGCTGTTCCTCCGCTGCTTTTGTCTCCACGGGCGAGA 1376
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
539 TTTATTTGGCGCGATTTTANGGCACTTGGATNTTTCGCTTGGCTTGGCTCCAGGGGCGAGA 480
QY 1377 CCTGACTACACTGCACTGACGAGGCTTATGGCTCGTGTCCGAGACCCCGACCCCT 1436
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
479 CCTTGATTACAA-TGCAGTGACGAGGCGCTATGGCTCGTGTCCGAGACCCCGNACCCCT 421
QY 1437 CCTTGAATCTGCTCCATCAATATGTGAGCCAGAGTATATCCAGCTGTTAATC-GATT 1495
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
420 CCTTGAATCTGCTCCATCAATATGTGAGCCAGAGTATATCCAGCTGTTAATCGATT 361
QY 1496 TTGGTGTCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAG 1555
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
360 TTGGTGTCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAG 301
QY 1556 GCATTGCTATGCTGTACAGGCGGAGCCACTCCAGGTCACCTTCTATCACAGTCCGTT 1615
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
300 GCATTGCTATGCTGTACAGGCGGAGCCACTCCAGGTCACCTTCTATCACAGTCCGTT 242
QY 1616 TAGTGTGCTCCGAGAGCTTGTGCG-AGGCTGGCGAGCCACCAAGCCATCAACAGCTGGAT 1674
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
241 TAGTGTGCTCCGAGAGCTTGTGCGAGGCTGGCGAGCCACCAAGCCATCAACAGCTGGAT 182
QY 1675 ATTCCTCCCATGTTGATTAGTACTACCAAAACCAACTGTAATCTTTCAGTCTCCCGAGG 1734
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 ATTCCTCCCATGTTGATTAGTACTACCAAAACCAACTGTAATCTTTCAGTCTCCCGAGG 122
QY 1735 ACCTTATGATGCTTCCGAAACCACTGGGAGTCAAGTGTGAGAGCATTTACAGCTT 1794
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 ACCTTATGATGCTTCCGAAACCACTGGGAGTCAAGTGTGAGAGCATTTACAGCTT 62
QY 1795 CATCCACTTACCTGGAGTGTCTCTCTGTTATTATCTCCACATAAAATTTCTCCAGAAA 1854
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 CATCCACTTACCTGGAGTGTCTCTCTGTTATTATCTCCACATAAAATTTCTCCAGAAA 2
QY 1855 T 1855
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 T 1
RESULT 10
W76256 446 bp mRNA linear EST 17-OCT-1996
LOCUS

```

```

DEFINITION
z66f07.r1 Soares fetal heart NBHL19W Homo sapiens cDNA clone
IMAGE:345637 5' similar to PIR:S37771 S37771 ankyrin, erythrocyte -
mouse ;, mRNA sequence.
W76256.1 GI:1386638
VERSION W76256.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 446)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 945 Std Error: 0.00
Seq primer: mob.REGA+ET
FEATURES
Location/Qualifiers
1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1271012"
/db_xref="taxon:9606"
/clone="IMAGE:345637"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBHL19W"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCGAAGTGGAGCGCGGCATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Patima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
BASE COUNT 78 a 131 c 115 g 122 t
ORIGIN
Query Match 23.7%; Score 445; DB 14; Length 446;
Best Local Similarity 100.0%; Pred. No. 2.6e-60;
Matches 445; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 973 GTTCTGGGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGAGCTGTTTGCAGTTC 1032
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
2 GTTCTGGGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGAGCTGTTTGCAGTTC 61
QY 1033 CTCCTTAGCCATGTGTGATGTTGACAGTGTGGATGTCAAGGCACAGACGCCACTTTTC 1092
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
62 CTCCTTAGCCATGTGTGATGTTGACAGTGTGGATGTCAAGGCACAGACGCCACTTTTC 121
QY 1093 ACTGCTGTGAGTATGAGGCGCATCTGGAGTGTGTACGTGTGCTTTTGAAGCTGTGCTCT 1152
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
122 ACTGCTGTGAGTATGAGGCGCATCTGGAGTGTGTACGTGTGCTTTTGAAGCTGTGCTCT 181
QY 1153 CCTGCTGTGAGTATGAGGCGCATCTGGAGTGTGTACGTGTGCTTTTGAAGCTGTGCTCT 1212
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
182 CCTGCTGTGAGTATGAGGCGCATCTGGAGTGTGTACGTGTGCTTTTGAAGCTGTGCTCT 241

```

QY 1213 GTTGCTATCTCGCAGGAGCTCTTAGACCATGTGTCAGAGCCCAACGCTAAAGCTAACTA 1272
 Db 242 GTTGCTATCTCGCAGGAGCTCTTAGACCATGTGTCAGAGCCCAACGCTAAAGCTAACTA 301
 QY 1273 CCAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCCCTCTATTGTCGCCGAGCTCTAC 1332
 Db 302 CCAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCCCTCTATTGTCGCCGAGCTCTAC 361
 QY 1333 GGCACCTGGACTGTTTCCGCCCTGCTTTTGTCTCCAGGGGACAGCCCTGACTACAACTGC 1392
 Db 362 GGCACCTGGACTGTTTCCGCCCTGCTTTTGTCTCCAGGGGACAGCCCTGACTACAACTGC 421
 QY 1393 ACTGACCGGCTATTGGCTCGTG 1417
 Db 422 ACTGACCGGCTATTGGCTCGTG 446

RESULT 11 W58499/c

LOCUS 464 bp mRNA linear EST 15-OCT-1996
 DEFINITION IMAGE:341192 3', mRNA sequence.

ACCESSION W58499.1 GI:1365358

VERSION EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished

TITLE JOURNAL COMMENT

Contact: Wilton RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
 This clone is available royalty-free through LLM; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 979 Std Error: 0.00
 Seq primer: mob.RSGA-ET

High quality sequence stop: 351.

FEATURES source

1. .464
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GBB:1266567"
 /db_xref="taxon:9606"
 /clone="IMAGE:341192"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal heart NDHL19W"
 /notes="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']. double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NDHL19W."

BASE COUNT
 ORIGIN

116 a 91 c 142 g 112 t 3 others

Query Match 21.5%; Score 404.6; DB 14; Length 464;
 Best Local Similarity 96.7%; Pred. No. 5.7e-54;
 Matches 444; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 1401 GGGCCATTGCTCTGTGCCAAGACCCCGCACCTCTCTTGAATCTGCTCCATCATAA 1460
 Db 459 GGCNTTATGGGCTCGTGTGCCAGGACCCCGNACCTCTCTGGAATCTGCTCCATCATAA 400
 QY 1461 TTGTGAGCCA-CAGTATATCCAGCTG--TTAATCGATTTTGGTGTAAATCTACCTTCC 1517
 Db 399 TTGTGAGCAGAGTATATCCAGCTGGTAAATCGATTTTGGTGTAAATCTACCTTCC 340
 QY 1518 ATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCATTTGCTGTCTACAGGC 1577
 Db 339 ATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCATTTGCTGTCTACAGGC 280
 QY 1578 CCGAGCCACTCCAGGTCACTTCTATCAGAGTCCGTTTAGTCTCCGAGAGCCTTGTG 1637
 Db 279 CCGAGCCACTCCAGGTCACTTCTATCAGAGTCCGTTTAGTCTCCGAGAGCCTTGTG 220
 QY 1638 CC-AGCTGGCCAGCACAAGCCATCAACACCTGGATTTCTCCCATGTTGATTAGCT 1696
 Db 219 CCNAGGCTGGCCAGCACAAGCCATCAACACCTGGATTTCTCCCATGTTGATTAGCT 160
 QY 1697 ACCTAAACACCAACTGTAATCTTGCAGTCTCCCGAGGAACCTTATGATCCCTCCGAAAC 1756
 Db 159 ACCTAAACACCAACTGTAATCTTGCAGTCTCCCGAGGAACCTTATGATCCCTCCGAAAC 100
 QY 1757 GACCTGGGAGCTCAGTAGCTGGAGAGCATTTACAGCTCATCCATCTTACTTGAGTGTCT 1816
 Db 99 CACCTGGGAGCTCAGTAGCTGGAGAGCATTTACAGCTCATCCATCTTACTTGAGTGTCT 40
 QY 1817 CTCCTGTATTATCTCCCAATAAATTTCTCCAGAAAT 1855
 Db 39 CTCCTGTATTATCTCCCAATAAATTTCTCCAGAAAT 1

RESULT 12 BI344684

LOCUS 373212 MARC 2P1G Sus scrofa cDNA 5', mRNA linear EST 30-JUL-2001
 DEFINITION BI344684
 ACCESSION BI344684
 VERSION BI344684.1 GI:15037973

KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J., Vallet, J., Wise, T., Rohrer, G.A., Perlea, G., Sultana, R., Quackenbush, J. and Keele, J.W.

REFERENCE AUTHORS

Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly
 Mamm. Genome 13 (8), 475-478 (2002)

TITLE JOURNAL MEDLINE PUBMED

COMMENT 22213789

12226715

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCAGTCAGCAGC

Plate: 119 row: K column: 23

Seq primer: ATTAGGTGACATATAG.

Location/Qualifiers

1. .579

Qy	268	GTGTGCCATGAGGTTTATATTCTGCTCAAGCTTTCTCAGCCAGCATTACCTGAAGAT	327
Db	241	GTGTGCCATGAGGTTTATATTCTGCTCAAGCTTTCTCAGCCAGCATTACCTGAAGAT	300
Qy	328	CTTTATGCTTTTCTTATATCCCAATCCTCAAAAGAGATGAGGGAAGTGGATGGAAA	387
Db	301	CTTTATGCTTTTCTTATATCCCAATCCTCAAAAGAGATGAGGGAAGTGGATGGAAA	360
Qy	388	CT 389	
Db	361	CT 362	

Search completed: December 1, 2003, 13:31:09
Job time : 3620 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 10:38:24 ; Search time 70 Seconds
(without alignments)
1655.223 Million cell updates/sec

Title: US-09-941-831A-20

Perfect score: 2343

Sequence: 1 MRESGWKLIDPISDFGRMGI.....QAINQLDIPPMISVLKQL 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	DB	ID	Description
1	1636	69.8	318	4 Q8N8F5	Q8N8F5 homo sapien
2	1435	61.2	308	11 Q9D738	Q9D738 mus musculus
3	675.5	28.8	704	4 Q96EFO	Q96EFO homo sapien
4	673	28.7	401	4 Q9Y3N3	Q9Y3N3 homo sapien
5	480.5	20.5	327	11 Q9D959	Q9D959 mus musculus
6	421	18.0	617	11 Q8VE11	Q8VE11 mus musculus
7	408	17.4	621	4 Q96F80	Q96F80 homo sapien
8	396	16.9	467	11 Q8K0A2	Q8K0A2 mus musculus
9	396	16.9	660	11 Q8K4J6	Q8K4J6 mus musculus
10	360	15.4	676	5 Q9W1Q6	Q9W1Q6 drosophila
11	360	15.4	761	5 Q8MLR7	Q8MLR7 drosophila
12	313.5	13.4	676	5 Q45553	Q45553 caenorhabdi
13	313.5	13.4	782	5 Q81412	Q81412 caenorhabdi
14	292.5	12.5	643	11 Q8VHA7	Q8VHA7 mus musculus
15	283	12.1	363	4 Q8NEC6	Q8NEC6 homo sapien
16	272.5	11.6	278	11 Q8VBX0	Q8VBX0 mus musculus

17	266.5	11.4	329	11 Q9D1A4	Q9D1A4 mus musculus
18	261.5	11.2	276	4 Q8N7B5	Q8N7B5 homo sapien
19	258	11.0	603	11 Q9QV59	Q9QV59 mus musculus
20	258	11.0	603	11 Q9D4L1	Q9D4L1 mus musculus
21	253.5	10.8	355	11 Q8CD00	Q8CD00 mus musculus
22	249.5	10.6	284	11 Q8CC66	Q8CC66 mus musculus
23	238.5	10.2	2027	13 Q8UW52	Q8UW52 fugu rubrip
24	237	10.1	1599	11 Q99NH0	Q99NH0 mus musculus
25	235	10.0	435	4 Q9HA95	Q9HA95 homo sapien
26	235	10.0	616	4 Q96GK0	Q96GK0 homo sapien
27	235	10.0	627	4 Q8WY90	Q8WY90 homo sapien
28	235	10.0	1486	4 Q8TEF1	Q8TEF1 homo sapien
29	235	10.0	2542	4 Q81WZ3	Q81WZ3 homo sapien
30	235	10.0	2617	4 Q81WZ2	Q81WZ2 homo sapien
31	233.5	10.0	525	11 Q9CWK6	Q9CWK6 mus musculus
32	233	9.9	627	4 Q96G77	Q96G77 homo sapien
33	231	9.9	833	4 Q96I86	Q96I86 homo sapien
34	231	9.9	1188	4 Q9H288	Q9H288 homo sapien
35	230.5	9.8	525	11 Q8R5B4	Q8R5B4 mus musculus
36	228	9.7	4001	5 Q8WR07	Q8WR07 drosophila
37	224.5	9.6	321	5 Q9VQ11	Q9VQ11 drosophila
38	224.5	9.6	619	5 Q9VM19	Q9VM19 drosophila
39	224.5	9.6	1719	4 Q13768	Q13768 homo sapien
40	224.5	9.6	1848	11 Q61302	Q61302 mus musculus
41	224.5	9.6	1856	4 Q99407	Q99407 homo sapien
42	224	9.6	323	11 Q9CQ31	Q9CQ31 mus musculus
43	223.5	9.5	1136	6 Q9N180	Q9N180 bos taurus
44	223	9.5	1181	5 Q9XZ37	Q9XZ37 drosophila
45	223	9.5	1181	5 Q9VBP3	Q9VBP3 drosophila

ALIGNMENTS

RESULT 1

Q8N8F5 PRELIMINARY; PRT; 318 AA.
 ID Q8N8F5
 AC Q8N8F5;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein FLJ39577.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RA Tachiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
 RA Isogai T.,
 RT "NEDO human cDNA sequencing project";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK096896; BAC04888.1;
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00023; ank; 5.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 5.
 DR PROSITE; PS50088; ANK_REPEAT; 3.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR Hypothetical protein; ANK repeat; Repeat.
 SQ SEQUENCE 318 AA; 34980 MW; C4D8F57AF22E629F CRC64;

Query Match 69.8%; Score 1636; DB 4; Length 318;
 Best Local Similarity 99.7%; Pred. No. 1.1e-137;
 Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 132 MRIVLQAKNMLDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYTLQQLRQERY 191
 DB 1 MRIVLQAKNMLDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYTLQQLRQERY 60
 QY 192 KRFINSRSGWGPVGTPLRLAASGYHLSCQLVLLAHGADVDSLDVKAQTPLFTAVSHGHL 251
 DB 61 KRFINSRSGWGPVGTPLRLAASGYHLSCQLVLLAHGADVDSLDVKAQTPLFTAVSHGHL 120
 QY 252 CVRVLLAAGASPGSGSYNNCSPLVTAARDGAVAILQELLHGAENAVKALPWNASNIAS 311
 DB 121 CVRVLLAAGASPGSGSYNNCSPLVTAARDGAVAILQELLHGAENAVKALPWNASNIAS 180
 QY 312 CSGPLYLAAYGYHLDCFRLLHLLHAGADPDYNTDQGLLARVPRTLEICLHNCPEYI 371
 DB 181 CSGPLYLAAYGYHLDCFRLLHLLHAGADPDYNTDQGLLARVPRTLEICLHNCPEYI 240
 QY 372 QLLIDFGANYLPSSLDLTSDQDKGIALLLQARATPRSLLSQVRLVRRALCQAGQPOA 431
 DB 241 QLLIDFGANYLPSSLDLTSDQDKGIALLLQARATPRSLLSQVRLVRRALCQAGQPOA 300
 QY 432 INQLDIPPMILSYLKHQL 449
 DB 301 INQLDIPPMILSYLKHQL 318

RESULT 2

Q9D738 PRELIMINARY; PRT; 308 AA.
 AC Q9D738;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE 2310036C05Rik protein (Ankyrin repeat domain-containing SOCS box
 DE protein Asb-12).
 GN 2310036C05Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giasi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzaerelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Hayashizaki Y.,
 RA Yashaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690 (2001).
 RN [2]
 SEQUENCE FROM N.A.
 RA Kile B.T., Nicola N.A.;
 RT "SOCS box proteins";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK009644; BAB26410.1; -.
 DR EMBL; AF403040; AAL57359.1; -.
 DR MGD; MGI:1917642; 2310036C05Rik.
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00023; ank; 5.

DR PRINTS; PRO1415; ANKYRIN.
 DR SMART; SM00248; ANK; 5.
 DR PROSITE; PS00088; ANK_REPEAT; 3.
 DR PROSITE; PS0297; ANK_REPEAT; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 308 AA; 33911 MW; 49824BE6616DB961 CRC64;

Query Match 61.2%; Score 1435; DB 11; Length 308;
 Best Local Similarity 88.6%; Pred. No. 9.7e-120; Indels 0; Gaps 0;
 Matches 273; Conservative 17; Mismatches 18;

QY 141 MNLMIDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYTLQQLRQERYKRFINSRSG 200
 DB 1 MNLMIDITKIFSLQPDKEEDTGTGKQALNQAVYDNDSTYTLQQLRQERYKRFINSRSG 60
 QY 201 MCVPGTPLRLAASGYHLSCQLVLLAHGADVDSLDVKAQTPLFTAVSHGHLDCVRVLL 260
 DB 61 MCVPGTPLRLAASGYHLSCQLVLLAHGADVDSLDVKAQTPLFTAVSHGHLDCVRVLL 120
 QY 261 ASPGSGSYNNCSPLVTAARDGAVAILQELLHGAENAVKALPWNASNIASCGPLYLAA 320
 DB 121 ASPGSGSYNNCSPLVTAARDGAVAILQELLHGAENAVKALPWNASNIASCGPLYLAA 180
 QY 321 VYGHLDLCPRLHLLHAGADPDYNTDQGLLARVPRTLEICLHNCPEYIQLLIDFGAN 380
 DB 181 VYGHLDLCPRLHLLHAGADPDYNTDQGLLARVPRTLEICLHNCPEYIQLLIDFGAN 240
 QY 381 IYLPSSLDLTSDQDKGIALLLQARATPRSLLSQVRLVRRALCQAGQPOAQNQLDIPPM 440
 DB 241 IYLPSSLDPDPTSDQDKGIALLLQARATPRSLLSQVRLVRRALCQAGQPOAQNQLDIPPM 300
 QY 441 LISYLKHQ 448
 DB 301 LISYLKHQ 308

RESULT 3

Q96EF0 PRELIMINARY; PRT; 704 AA.
 ID Q96EF0
 AC Q96EF0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Similar to hypothetical protein FLJ20126.
 DE Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SEQUENCE FROM N.A.
 RP TISSUE=Ovary;
 RC Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC012399; AAH12399.1; -.
 DR Genew; HGNC:16825; MTMR8.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 KW Hydrolyase.
 SQ SEQUENCE 704 AA; 78918 MW; 515BE817C9AEA961 CRC64;

Query Match 28.8%; Score 675.5; DB 4; Length 704;
 Best Local Similarity 38.7%; Pred. No. 2.3e-51;
 Matches 186; Conservative 44; Mismatches 128; Indels 123; Gaps 18;

QY 1 MRESGKLIIDPISDFGRMGIPNRRNTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
 DB 122 MRESGKLIIDPISDFGRMGIPNRRNTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 181
 QY 61 FRSKERVPLSYLYKENNAIACRCQPLSGFTRCVDDLELLEAISTQTPGQFMVVDVT 120
 DB 182 FRSKERVPLSYLYKENNAIACRCQPLSGFTRCVDDLELLEAISTQTPGQFMVVDVT 241
 QY 121 RPKI-----WHFL-----VLIMRIVLQAKNMLDITKIFSLQ 154

```

|||||
242 REKLNAMANRAAGKGYENEDNYANIRFRPMGIENIHVMSRLQ-----KLEVECELKT 294
QY Q9Y3N3 PRELIMINARY; PRT; 401 AA.
AC Q9Y3N3;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE DJ710L4.2 (Similar to MYOTUBULARIN-related protein) (Fragment).
GN DJ710L4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chapman J.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034408; CAB38778.1; -.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hydrolyase.
FT NON_TER 1
FT TER 401
SQ SEQUENCE 401 AA; 45607 MW; 8044A4031B35298C CRC64;

Query Match 28.7%; Score 673; DB 4; Length 401;
Best Local Similarity 50.7%; Pred. No. 1.7e-51;
Matches 155; Conservative 28; Mismatches 59; Indels 64; Gaps 10;

QY 1 MRESGWKLIDPISDFGRMGIPNRNWTITDANRYEICSTYPPEIVVPKSVTLGTWVGSSK 60
DB |||||
73 MRESGWKLIDPISDFGRMGIPNRNWTITDANRYEICSTYPPEIVVPKSVTLGTWVGSSK 132
QY 61 PESKERVPLSYLYKENNAACRCQPSLGSFGFYTRCVDDLELLEASQTPNGSQFMVYVD 120
DB |||||
133 FSKKERVPLSYLYKENNAACRCQPSLGSFGFYTRCVDDLELLEASQTPNGSQFMVYVD 192
QY 121 RPKI-----WHFL-----VLIMRVLQAKMNLMDITKIFSLIQ 154
DB |||||
193 REKLNAMANRAAGKGYENEDNYANIRFRPMGIENIHVMSRLQ-----KLEVECELKT 245
QY 155 PKKEEDTDEKQALN--QAVYDNDSTYTLDLLQROERYKRFRINSRSGWVGPGTPLRLAA 212
DB |||||
246 PTMSEFLSGLESSGWLRIKAIMDAGIF-ITRAVKVEKASVLVHCSGDGMDRTAQCVSVAS 304

```

```

213 -----SYGHLSCLOVLL-----AHGADVDSLVDVKAQTPLFTAVSHGHLD 252
DB |||||
305 ILLDPFYRTFKGLMILIEKEMWISMGHKFSQRCGH-LDGDSEK-----SPIFTQ----FLDC 356
QY 253 VRVLE 258
DB |||||
357 IWQLME 362

RESULT 5
Q9D9S9 PRELIMINARY; PRT; 327 AA.
AC Q9D9S9;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE 1700023008Rik protein.
GN ASB1 OR 1700029008RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Mateu Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Butt C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 403:685-690(2000).
DR EMBL; AK006515; BAB24628.1; -.
DR HSSP; Q00420; IAWC.
DR MGD; MGI:1929735; Asb1.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 5.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 327 AA; 36281 MW; CB0649F3D080910C CRC64;

Query Match 20.5%; Score 480.5; DB 11; Length 327;
Best Local Similarity 40.7%; Pred. No. 2e-34;
Matches 111; Conservative 51; Mismatches 90; Indels 21; Gaps 7;

QY 170 LNAQVYDNDSTYTLDLLQROERYKRFRINSRSGW--GVFGTPLRLAASYGHLSCLOVLLAH 286
DB |||||
34 LHDAAYVGDQLQTLRNLLQEEYSRYSRINSEKSWCCWGLPCTPURIATATAGHCNVDFLIRK 93
QY 227 GADVDSLVDVKAQTPLFTAVSHGHLDVRLVLEAGASPGSIVNNCSPLVLTAAARDGAVAIL 286
DB |||||
94 GAEDVLDVKGQALYVAVVNGHLESTELLEAGADPNGRHRSRTPYVHASRVGDDLL 153
QY 287 QELLDHGAENYAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLHLHGADP 338
DB |||||
154 KALIRYAGADVNVNHLTPDTRPPFSRRRLTSLVLC--PLYISAAYHNLQCFRLLQAGANP 211

```

```

QY 339 DYNCTDQGLLARVPR--PRTLLEICLHNCCEPEYIOLLIDFGANIYL---PSLSLDTLSQ 393
Db 212 DFNCNGPNTQBYRSGPCVMDVLRHGCCEAAVSLVFEFGANLNLVKWESLGPARGR 271
QY 394 ---DDKGIALQLQARATPRSLISQVRLVRRAL 423
Db 272 RKMDPEALQVKEAREIPRTLISLCRVAVRRAL 304

RESULT 6
Q8VE11 PRELIMINARY; PRT; 617 AA.
AC Q8VE11;
DT 01-WAR-2002 (TrEMBLrel. 20, Created)
DT 01-WAR-2003 (TrEMBLrel. 20, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (MYOTUBULARIN related protein 6 homolog).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; BC020019; AAH20019.1; -.
DR EMBL; AK076218; BAC36259.1; -.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hypothetical protein; Hydrolase.
SQ SEQUENCE 617 AA; 70932 MW; EE699C578602A013 CRC64;

Query Match 18.0%; Score 421; DB 11; Length 617;
Best Local Similarity 59.3%; Pred. No. 1e-28;
Matches 73; Conservative 25; Mismatches 25; Indels 0; Gaps 0;

QY 2 RESGKWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 121 RRGWQLIDLAEEYKRMGVNPNWQLSDANREYKVCETYPRELYVPTASRPVIVGSSNF 180
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 181 RSKGRPLVLSYCRQGTAAICRCSQPLSGFSARCLEDEHLLQAISKANPNRYMYVDTR 240

Query Match 16.9%; Score 396; DB 11; Length 467;
Best Local Similarity 55.3%; Pred. No. 1.2e-26;
Matches 68; Conservative 25; Mismatches 30; Indels 0; Gaps 0;

QY 2 RESGKWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 123 REQGLVLDLSEYKRMGLPDNYWQLSDNRYVCDSYFTLYVPRSATAHIIIVGSSKF 182
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 183 RSRRPPLSYKCKSHASICKSSQPLSGFSARCLEDEHLLQAISKANPNRYMYVDTR 242
QY 122 PKI 124
Db 243 PKL 245

RESULT 7
Q96P80 PRELIMINARY; PRT; 621 AA.
AC Q96P80;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Myotubularin related protein 6.
GN MTMR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Hong W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF408619; AAL01037.1; -.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hydrolase.
SQ SEQUENCE 621 AA; 71896 MW; A7346D3AE82911A7 CRC64;

Query Match 17.4%; Score 408; DB 4; Length 621;
Best Local Similarity 58.5%; Pred. No. 1.5e-27;
Matches 72; Conservative 25; Mismatches 26; Indels 0; Gaps 0;

QY 2 RESGKWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 121 RLQHQQLIDLAEEYKRMGVNPNWQLSDANREYKVCETYPRELYVPTASRPVIVGSSKF 180
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 181 RSKGRFPVLSYHYHDKAICRCSQPLSGFSARCLEDEHLLQAISKANPNRYMYVDTR 240
QY 122 PKI 124
Db 241 PKL 243

RESULT 8
Q8K0A2 PRELIMINARY; PRT; 467 AA.
AC Q8K0A2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to myotubularin related protein 6.
GN MTMR7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032254; AAH32254.1; -.
DR MGI; MGI:1891693; Mtnr7.
DR InterPro; IPR003595; PTPC_motif.
DR InterPro; IPR000387; TYR_phosphatase.
DR SMART; SM00404; PTPC motif; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
SQ SEQUENCE 467 AA; 53927 MW; F756FB90E316A9A5 CRC64;

Query Match 16.9%; Score 396; DB 11; Length 467;
Best Local Similarity 55.3%; Pred. No. 1.2e-26;
Matches 68; Conservative 25; Mismatches 30; Indels 0; Gaps 0;

QY 2 RESGKWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 123 REQGLVLDLSEYKRMGLPDNYWQLSDNRYVCDSYFTLYVPRSATAHIIIVGSSKF 182
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 183 RSRRPPLSYKCKSHASICKSSQPLSGFSARCLEDEHLLQAISKANPNRYMYVDTR 242
QY 122 PKI 124
Db 243 PKL 245

RESULT 9
Q8C4J6 PRELIMINARY; PRT; 660 AA.
ID Q8C4J6
AC Q8C4J6;

```

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Myotubularin related protein 7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL; AK081973; BAC38383.1; --
 SQ SEQUENCE 660 AA; 75608 MW; 57CA1A447192A52B CRC64;
 Query Match 16.9%; Score 396; DB 11; Length 660;
 Best Local Similarity 55.3%; Pred. No. 28-26;
 Matches 68; Conservative 25; Mismatches 30; Indels 0; Gaps 0;
 QY 2 RESGWKLIDPISDFGRMGIPNRRNWTITDANRYEICTYPPETVVPKSVTLGTWVGSSKF 61
 DB 123 RQGMLLVLSBEYKEMGLPDNYWQSLDVRDYVCDSTELYVPRSAHIIVGSSKF 182
 QY 62 RSKERVPLVSLYKKNNAICRCSPLSGFYTRCVDDELLEAISOTNPGSQPMYVVDTR 121
 DB 183 RGRRRPPALSYKDSHASICRQSPLSGFSARCLEDQMLQAIKANKPGSDFIYVVDTR 242
 QY 122 PKI 124
 DB 243 PKL 245
 RESULT 10
 Q9W1Q6 PRELIMINARY; PRT; 676 AA.
 AC Q9W1Q6; Q9V3A1;
 DT 01-WAY-2000 (TrEMBLrel. 13, Created)
 DT 01-WAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE BCDNA:GH04637 protein.
 GN BCDNA:GH04637 OR C3530 OR C3530 OR C3530 OR C3530
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam M.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zeng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC STRAIN=BERKELEY; TISSUE=Head;
 RX MEDLINE=20196012; PubMed=10731138;
 RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
 RA Stapleton M., Harvey D.A.;
 RT "A Drosophila complementary DNA resource";
 RL Science 287:2222-2224 (2000).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL; AE003460; AAF46997.1; --
 DR EMBL; AE003460; AAF46999.1; --
 DR EMBL; AF181629; AAD55415.1; --
 DR Flybase; FBgn0028497; BCDNA:GH04637.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR000387; TYR phosphatase.
 DR InterPro; IPR000306; Znf_FYVE.
 DR Pfam; PF01363; FYVE; 1.
 DR SMART; SM00064; FYVE; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00383; TYR PHOSPHATASE_1; 1.
 DR PROSITE; PS00566; TYR PHOSPHATASE_2; 1.
 DR PROSITE; PS0178; ZF_FYVE; 1.
 KW Hypothetical protein, Alternative splicing; Hydrolase.
 FT VARSPPLIC 1 37 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 676 AA; 77223 MW; 988BE4C052754B43 CRC64;
 Query Match 15.4%; Score 360; DB 5; Length 676;
 Best Local Similarity 26.9%; Pred. No. 3.4e-23;
 Matches 119; Conservative 62; Mismatches 159; Indels 102; Gaps 15;
 QY 2 RESGWKLIDPISDFGRMGIPNRRNWTITDANRYEICTYPPETVVPKSVTLGTWVGSSKF 61
 DB 108 KNAGWDYFKLEAEFKHMLVNEAWTLCSSNNKEYELCDYPRQIYVPEATLMLISSRF 167
 QY 62 RSKERVPLVSLYKKNNAICRCSPLSGFYTRCVDDELLEAISOTNPGSQPMYVVDTR 121
 DB 168 RSKRGRLPVLTYLH-NNKASICRCSPLSGFSARCLEDQMLQAIKANKNTNDIYVVDTR 226
 QY 122 PKI-----WHFLVLMRIVLQAKNMLMDITKIFSLQPDKEE 159
 DB 227 PRINAMNPAAGKGYNEAFYENIKHFILG-IENIHVQRASLQ-----KYLEACEQK 277
 QY 160 EDTDTEEKQALN-----QAVYNDSTYTLQDLQERYKRFINSRSGVPG-----T 206
 DB 278 SPTMSAFINALESSGWLKHRSILDTSSFIANAV--DKGVSVVHCHSDGMDRTAQVCSLA 335
 QY 207 PRLAASYGHLSCQLV-----LAHCADVD-----SLDVKAQTPLFTAVSHGHDCV-- 253
 DB 336 QLMNLPYRTIKGFQALIEKDWLAFGHKFSRCGHITQTDAREVSPFTQ-----FLDCTWQ 391
 QY 254 -----RVLL-----EAGSPGSGSYNNCSFVLTAARDGAVAILQELLDHGAEA 296

```

Db 392 LMSORSEAFENRFLILLHDHVHSCQFGTFVGNCE-----KORLDLKLAE 437
Qy 297 NVKAKLPVWASNTASCGPLYL-----AAVYGHLDCEFRLLLLHGADPPDYNCTDQGLLARVP 352
Db 438 RTFSLMGYMANHLEINYPINPKNVDEAIKANLAPQICKWRGM---YSRFSGIHPREP 494
Qy 353 RPTLLICLHNCPEPIYQILL 374
Db 495 LGDVLDSKHCNSLESDHVQHL 516

RESULT 11
Q8MLR7
ID Q8MLR7 PRELIMINARY; PRT; 761 AA.
AC Q8MLR7;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE CG3530-PA.
GN BCDNA:GH04637 OR CG3530 OR CG18093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophiliidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley; PubMed=107311132;
RX MEDLINE=20196006;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.G., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Franknoch C., Baldwin E.M.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

```

```

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreesnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Klomp W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003460; AAF46998.2; -.
DR FlyBase; FBGN0028497; BCDNA:GH04637.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003595; PTPC motif.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00404; PTPC motif; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00556; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
SQ SEQUENCE 761 AA; 87047 MW; 11BED6502933A860 CRC64;
Query Match 15.4%; Score 360; DB 5; Length 761;
Best Local Similarity 26.9%; Pred. No. 4e-23;
Matches 119; Conservative 62; Mismatches 159; Indels 102; Gaps 15;
Qy 2 RESGKLLDIPIDGFRGMIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 193 KNAGWDYFKLEAFKHLVLPNEAWTICSNWEKELCDYTPROIYVPKENTLMLISSRP 252
Qy 62 RSKRPVPLSVLYKNNAAICRCSQPLSGFYTRCVDDLELLLEAISQTNFGSOFMYVVDTR 121
Db 253 RSKGRPLVLYLH-NKASICRCSQPLSGFSARCLEDQMLEAIRKNTSDYMYVVDTR 311
Qy 122 PKI-----WHFLVLIMRIVLQAVKQNLMDITKIFSLQPDKEE 159
Db 312 PRINAMANAAGKGYENAEFYENIKFHLG-IENIHVQRASLQ-----KVLEACEQK 362
Qy 160 EDTDTEERQALN-----QAVYNDNSVTLDQLLRQERYKRFINSRSGVPG-----T 206
Db 363 SPTMSAFNALLESSGWLKHIRSLDTSSTFIANAV--DKGVSVVVHCSGDGWRDTAQVCSLA 420
Qy 207 PLRLAASYGHLSCLQVL-----LAHGADV-----SLDKVAQTPLFTAVSHGHDCV--- 253
Db 421 QLMNLNPPYRTIKGFQALLIEKQWLAFGHKFSERCGHQTDAREVSPFTQ-----FLDCTWQ 476
Qy 254 -----RVLL-----EAGASPGSGSYNNCSPLVTAARDGAVAILQELLHGARA 296
Db 477 LMSORSEAFENRFLILLHDHVHSCQFGTFVGNCE-----KORLDLKLAE 522

```



```

QY 297 NVKAKLPVWASNTASCGPLYL-----AAYVGHLDGFRLLHGGADPDYNTDQGLLARPV 352
DB 523 RTFSLGWYMANHLNVEINPLKPNVDIAKANLAPQCIKFWGCM--YSRFESGIHPREP 579
QY 353 RPTLLIEICLHNCPEFYIQLL 374
DB 580 LGDVLDDSKHCNSLEDHVQHL 601

RESULT 12
ID Q45553 PRELIMINARY; PRT; 676 AA.
AC Q45553;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE F53A2.8 protein.
GN F53A2.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RA Lloyd C.R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81546; CAB04456.1; -.
KW Hypothetical protein.
SQ SEQUENCE 782 AA; 88303 MW; 013FALL103B50E48C CRC64;

Query Match 13.4%; Score 313.5; DB 5; Length 782;
Best Local Similarity 35.3%; Pred. No. 4.8e-19;
Matches 59; Conservative 37; Mismatches 48; Indels 23; Gaps 2;

QY 5 GWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPEIVVPKSVTLGTVVGSSKFRSK 64
DB 125 GWRRLDWNSEMTROGITSQWTSNINEGYTICDTPNKLWPTTAASTSVLLGSCCKFRSR 290
QY 65 ERVPVLSYLKKNNAICRCQPSLGSFYTRCVDDLELLLEAISQTPGSMYVVDTRPKI 124
DB 185 GRPLVLTYPHQTEAALCRCAQPLTGFSARCVEDEKLMELVGKANTNSDNLFLVDTRPRV 244
QY 125 -----WHFLVLMIRIVLQAKNLMND-ITK 148
DB 245 NAMVNVKQKGFEDERNYSNMRFFPDIENTIHVMRASQARLLDDAVTK 291

Query Match 13.4%; Score 313.5; DB 5; Length 676;
Best Local Similarity 35.3%; Pred. No. 4.8e-19;
Matches 59; Conservative 37; Mismatches 48; Indels 23; Gaps 2;

QY 5 GWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPEIVVPKSVTLGTVVGSSKFRSK 64
DB 125 GWRRLDWNSEMTROGITSQWTSNINEGYTICDTPNKLWPTTAASTSVLLGSCCKFRSR 184
QY 65 ERVPVLSYLKKNNAICRCQPSLGSFYTRCVDDLELLLEAISQTPGSMYVVDTRPKI 124
DB 185 GRPLVLTYPHQTEAALCRCAQPLTGFSARCVEDEKLMELVGKANTNSDNLFLVDTRPRV 244
QY 125 -----WHFLVLMIRIVLQAKNLMND-ITK 148
DB 245 NAMVNVKQKGFEDERNYSNMRFFPDIENTIHVMRASQARLLDDAVTK 291

RESULT 13
ID Q81412 PRELIMINARY; PRT; 782 AA.
AC Q81412;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein F53A2.8b.
GN F53A2.8b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;

```

```

RN [1]
SEQUENCE FROM N.A.
RA Lloyd C.R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA none;
RX MEDLINE=99069613; PubMed=9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81546; CAB045596.2; -.
KW Hypothetical protein.
SQ SEQUENCE 782 AA; 88303 MW; 013FALL103B50E48C CRC64;

Query Match 13.4%; Score 313.5; DB 5; Length 782;
Best Local Similarity 35.3%; Pred. No. 5.9e-19;
Matches 59; Conservative 37; Mismatches 48; Indels 23; Gaps 2;

QY 5 GWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPEIVVPKSVTLGTVVGSSKFRSK 64
DB 231 GWRRLDWNSEMTROGITSQWTSNINEGYTICDTPNKLWPTTAASTSVLLGSCCKFRSR 290
QY 65 ERVPVLSYLKKNNAICRCQPSLGSFYTRCVDDLELLLEAISQTPGSMYVVDTRPKI 124
DB 291 GRPLVLTYPHQTEAALCRCAQPLTGFSARCVEDEKLMELVGKANTNSDNLFLVDTRPRV 350
QY 125 -----WHFLVLMIRIVLQAKNLMND-ITK 148
DB 351 NAMVNVKQKGFEDERNYSNMRFFPDIENTIHVMRASQARLLDDAVTK 397

RESULT 14
Q8VHA7
ID Q8VHA7 PRELIMINARY; PRT; 643 AA.
AC Q8VHA7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Myotubularin-related protein 2.
GN MTR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA Bolino A., Marigo V., Loader J., Romio L., Leoni A., Di Duca M.,
Cinti R., Feltri M.L., Wrabetz L., Ravazzolo R., Monaco A.P.;
RT "Molecular characterization and expression analysis of Mtnr2, mouse
homolog of MTR2, the myotubularin-related 2 gene, mutated in CMT4B.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY055832; AAL14198.1; -.
DR MGD; MGI:1924366; Mtnr2.
DR InterPro; IPR004182; GRAM dom.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF02893; GRAM; 1.
DR SMART; SM00568; GRAM; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
KW Hydrolase.
SQ SEQUENCE 643 AA; 73255 MW; E577B41175C2792F CRC64;

Query Match 12.5%; Score 292.5; DB 11; Length 643;
Best Local Similarity 25.1%; Pred. No. 3.4e-17;
Matches 99; Conservative 56; Mismatches 120; Indels 119; Gaps 17;

QY 3 ESWGKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPEIVVPKSVTLGTVVGSSKFR 62
DB 203 ESWGKLIDPILLYRRQGIPIESWRITKINERVELCDTYPALLIVVPANIPDELUKRVASFR 262
QY 63 SKERPVLVLSYLKKNNAICRCQPSLGSFY-TRCVDDLELLLEAISQTPGSMYVVDTR 121

```

263	DB	SRGRIPVLSWHPESQATVTRCSQBMVCVSGKRSKEDEKYLQALIMDSNAQSHKIFIPDAR	322
122	QY	PKI-----WHFLVLMIRIVLQAKNLM--DITKIFSLQPDKEE----	159
323	DB	PSVNAVANKAGGYESEDAYQNAELVF--LDIHNTHVMRESLRKLKEIVTYTIBETHW	379
160	QY	---EDT-----DTEKQALNAQAVYDNDSY-----TLDQLLRQERYK	192
380	DB	LSNLESTHWLEHILILAGALRIADKVESGKTSVVVHCSDGWDRTQALTSLAMLMIDGY	439
193	QY	RFINS-----RSGWGVPGTPIRLAASYGHLSCLQVLLAHG---ADVDSLDVKAQTPLFT	243
440	DB	RTIRFGEVLVEKEW-----LSFGHR--FOLRVGHGDKNHADAD---RXPVPL	481
244	QY	AVSHGHLDCVRVILLEGASPGSGSYNNCSVPLTAARDGAVAILQELLDH-----	292
482	DB	Q-----FIDCVQWQMT--OFFTAFENE-----YFLITIDHLHLSCLFGTFLC	522
293	QY	-----GAEANVKAKLPIWA-----SNIASCSGPLY	317
523	DB	NSEQQRGENLPKKTIVSLWSYINSQLEDFNPLY	556

RESULT 15

Q8NEC6		PRELIMINARY;	PRT; 363 AA.
ID	Q8NEC6		
AC	Q8NEC6;		
DT	01-OCT-2002	(TREMBLrel. 22, Created)	
DT	01-OCT-2002	(TREMBLrel. 22, Last sequence update)	
DE	01-MAR-2003	(TREMBLrel. 23, Last annotation update)	
DE	Similar to myotubularin related protein 1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
NCBI_TaxID=9606;			
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	TISSUE=Testis;		
RC	Strausberg R.;		
RA	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.		
RL	EMBL, BC032947; AAH32947.1; -		
DR	InterPro; IPR004182; GRAM_dom.		
DR	Ffam; PF02893; GRAM; 1.		
DR	SMART; SMO0568; GRAM; 1.		
DR	SEQUENCE 363 AA; 39822 MW; 9AE00536BCEAF0811 CRC64;		
SQ			

Search completed: December 1, 2003, 13:45:30
Job time : 76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2003, 09:17:14 ; Search time 6732 Seconds
(without alignments)
11412.400 Million cell updates/sec

Title: US-09-941-831A-6

Perfect score: 1878

Sequence: 1 catgattacgccagcttgg.....taaaaaaaaaaaaaaaaa 1878

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1119.4	59.6	1269	9	AK096896	AK096896 Homo sapi
2	1117.8	59.5	1222	6	AX418418	AX418418 Sequence
3	931	49.6	1069	6	AX418416	AX418416 Sequence
c	851.4	45.3	104534	9	AF403030	AF403030 Homo sapi
	729.2	38.8	1262	10	AF403040	AF403040 Mus muscu
	710.2	37.8	2653	9	BC012399	BC012399 Homo sapi
	703.4	37.5	1345	6	AX406008	AX406008 Sequence
c	628.6	33.5	191093	10	AL671765	AL671765 Mouse DNA
	612	32.6	194079	2	AC109692	AC109692 Rattus no
	612	32.6	225982	2	AC108644	AC108644 Rattus no
	290.6	15.5	240816	5	AL844570	AL844570 Zebrafish
13	254.2	13.5	2354	5	BC053411	BC053411 Danio rer
14	241.8	12.9	2353	6	AX086038	AX086038 Sequence
15	238.2	12.7	2754	5	BC044359	BC044359 Danio rer
16	233.6	12.4	2817	10	BC032254	BC032254 Mus muscu
17	228	12.1	3065	10	BC020019	BC020019 Mus muscu
18	224	11.9	1866	9	AF406619	AF406619 Homo sapi
19	224	11.9	2337	9	AK093237	AK093237 Homo sapi
20	224	11.9	4069	9	BC040012	BC040012 Homo sapi
21	220.8	11.8	4131	9	HS803324	HS803324 Homo sapi
22	204.4	10.9	1997	9	AK000133	AK000133 Homo sapi
23	181.4	9.7	3059	3	BT003304	BT003304 Drosophila
c	169	9.0	129849	9	HS71014	HS71014 Human DNA
	149.6	8.0	3209	3	AF181629	AF181629 Drosophila
	135.8	7.2	2482	6	AX398878	AX398878 Sequence
	109.6	5.8	1008	10	AF153352	AF153352 Mus muscu
28	109.6	5.8	1312	10	BC049699	BC049699 Mus muscu
29	109.6	5.8	5892	10	AK122456	AK122456 Mus muscu
30	108.8	5.8	2019	6	AR262497	AR262497 Sequence
31	108.8	5.8	2019	6	AX022304	AX022304 Sequence
32	108.8	5.8	2019	6	AX030771	AX030771 Sequence
33	108.8	5.8	2019	6	BD007458	BD007458 Remedies
34	103.6	5.5	3382	9	AF072928	AF072928 Homo sapi
35	100.8	5.4	1008	9	AF156777	AF156777 Homo sapi
36	100.8	5.4	1340	9	BC014528	BC014528 Homo sapi
37	100.2	5.3	1648	9	AF073482	AF073482 Homo sapi
38	93.6	5.0	6518	9	AB032972	AB032972 Homo sapi
39	92	4.9	190461	2	AC109199	AC109199 Mus muscu
40	92	4.9	228645	2	AC094273	AC094273 Rattus no
41	83	4.4	42724	3	CEP53A2	281546 Caenorhabdi
c	77.4	4.1	93778	9	AC016999	AC016999 Homo sapi
	77.4	4.1	172990	2	AC011973	AC011973 Homo sapi
	76.4	4.1	2397	9	BC052990	BC052990 Homo sapi
	76.4	4.1	3158	9	BC040432	BC040432 Homo sapi

ALIGNMENTS

RESULT 1
AK096896
LOCUS
DEFINITION Homo sapiens cDNA FLJ39577 fis, clone SKMUS200290, weakly similar
to Homo sapiens ASB-1 protein mRNA.
ACCESSION AK096896
VERSION AK096896.1 GI:21756490
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,
Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R.,
1
PRI 15-JUL-2002

/db_xref="taxon:32644"
/note="NOVX nucleic acid with homology to ankyrin
repeat-containing- (ASB-1) like protein"
BASE COUNT 276 a 352 c 278 g 316 t
ORIGIN

Query Match 59.5%; Score 1117.8; DB 6; Length 1222;
Best Local Similarity 99.8%; Pred. No. 4.3e-308;
Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 730 CCAAGATCTGCATTCCTTGTCTCATTAATGAGATAGTTCTCCAATTAGCCGAAGATG 789
DB |||||
102 CCACAGATCTGGCATTCCTTGTGCTCAATGAGATAGTTCTCCAATTAGCCGAAGATG 161
QY 790 AACCTCATGGACATCACCAAGATCTTCCCTCTCGCAGCCGACAAAGGAGGAGGAGAC 849
DB |||||
162 AACCTCATGGACATCACCAAGATCTTCTCCCTCTCGCAGCCGACAAAGGAGGAGGAGAC 221
QY 850 ACTGACACAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 909
DB |||||
222 ACTGACACAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 281
QY 910 TTGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 969
DB |||||
282 TTGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 341
QY 970 GGTGTTCTCGGACACCTTGGCTGGCTGCTCTTATGGCCACTTGAGCTGTTTGCAA 1029
DB |||||
342 GGTGTTCTCGGACACCTTGGCTGGCTGCTCTTATGGCCACTTGAGCTGTTTGCAA 401
QY 1030 GTCTCTTAGCCCATGCTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCACTT 1089
DB |||||
402 GTCTCTTAGCCCATGCTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCACTT 461
QY 1090 TTCACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1149
DB |||||
462 TTCACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 521
QY 1150 TCTCTGT 1209
DB |||||
522 TCTCTGT 581
QY 1210 GCTGTTGTCTATCTGACGAGCTCTTACACCATGCTGACAGGACCAAGCTCAAGCTAAA 1269
DB |||||
582 GCTGTTGTCTATCTGACGAGCTCTTACACCATGCTGACAGGACCAAGCTCAAGCTAAA 641
QY 1270 CTACAGCTCTGGGCACTCAAACTAGCTTCATGTTCTGGCCCTCTCTATTTGGCCGAGTC 1329
DB |||||
642 CTACAGCTCTGGGCACTCAAACTAGCTTCATGTTCTGGCCCTCTCTATTTGGCCGAGTC 701
QY 1330 TACGGGCACTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGACAGCCCTGACTACAC 1389
DB |||||
702 TACGGGCACTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGACAGCCCTGACTACAC 761
QY 1390 TGCACCTGACAGGCGCTATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1449
DB |||||
762 TGCACCTGACAGGCGCTATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 821
QY 1450 CTCCATCATTAATTTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTCTAATATC 1509
DB |||||
822 CTCCATCATTAATTTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTCTAATATC 881
QY 1510 TACCTTCATCTCTCTCCCTGACCTGACCTCACAAGATGATAAGGCAATTCATTTGCTG 1569
DB |||||
882 TACCTTCATCTCTCTCCCTGACCTGACCTCACAAGATGATAAGGCAATTCATTTGCTG 941
QY 1570 CTACAGGCGGAGGCACTCCAGCTGCTCTCTATACAGGTCCTGTTAGTGTGCTGCGCAGA 1629
DB |||||
942 CTACAGGCGGAGGCACTCCAGCTGCTCTCTATACAGGTCCTGTTAGTGTGCTGCGCAGA 1001
QY 1630 GCTTTGTGCGGCTGGCCAGGCAACAGGCACTCAACAGCTGGATATCTCTCCCATGTTG 1689
DB |||||
1002 GCTTTGTGCGGCTGGCCAGGCAACAGGCACTCAACAGCTGGATATCTCTCCCATGTTG 1061

QY 1690 ATTAGCTACCTAAACACCAACTGTATCTTGTGAGCTCTCCCGAGGAACTTATGATGCTTC 1749
DB |||||
1062 ATTAGCTACCTAAACACCAACTGTATCTTGTGAGCTCTCCCGAGGAACTTATGATGCTTC 1121
QY 1750 CGAAAACCACTGGGAGTCTACGTAGCTGGAGAGCATTTACAGCTCTACCTACCTACCTGG 1809
DB |||||
1122 CGAAAACCACTGGGAGTCTACGTAGCTGGAGAGCATTTACAGCTCTACCTACCTACCTGG 1181
QY 1810 AGCTGCTCTCTGTATTTATCTCCACAAATAAAATTTCTCCAG 1850
DB |||||
1182 AGCTGCTCTCTGTATTTATCTCCACAAATAAAATTTCTCCAG 1222

RESULT 3
AX418416
LOCUS AX418416 1069 bp DNA linear PAT 18-JUN-2002
DEFINITION Sequence 11 from Patent WO0206329.
ACCESSION AX418416
VERSION AX418416.1 GI:21523341
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 unclassified.
AUTHORS Rastelli,L., Shinkets,R.A., Zerhusen,B., Malyankar,U.M. and Padigaru,M.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0206329-A 11 24-JAN-2002;
Curagen Corporation (US)
FEATURES
1. 1069 Location/Qualifiers
source
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="NOVX nucleic acid with homology to ankyrin repeat-containing- (ASB-1) like protein"
BASE COUNT 237 a 298 c 257 g 277 t
ORIGIN

Query Match 49.6%; Score 931; DB 6; Length 1069;
Best Local Similarity 97.6%; Pred. No. 1.2e-254;
Matches 965; Conservative 0; Mismatches 0; Indels 24; Gaps 1;

QY 760 ATGAGATAGTCTCCAAATTAGCCAGATGAACTCTCAGGACATCACCAGATCTCTCC 819
DB |||||
81 ATGAGATAGTCTCCAAATTAGCCAGATGAACTCTCAGGACATCACCAGATCTCTCC 140
QY 820 CTCTGCGCCCGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 879
DB |||||
141 CTCTGCGCCCGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 200
QY 880 CAAGCAGTGTATGACCAAGCTCTTATCTTTGGACCAAGCTTTTGGCCAGAGGAGGAGGAG 939
DB |||||
201 CAAGCAGTGTATGACCAAGCTCTTATCTTTGGACCAAGCTTTTGGCCAGAGGAGGAGGAG 260
QY 940 AAACGTTTCATCAACAGCAGGAGTGGCTGGGGTGTCTTGGACACCCCTTGGCTTGGCT 999
DB |||||
261 AAACGTTTCATCAACAGCAGGAGTGGCTGGGGTGTCTTGGACACCCCTTGGCTTGGCT 320
QY 1000 GCTTCTTATGCGCACTTGCAGCTGTTTGAAGTCTCTTAGCCCATGGTGTCTGATGTTGAC 1059
DB |||||
321 GCTTCTTATGCGCACTTGCAGCTGTTTGAAGTCTCTTAGCCCATGGTGTCTGATGTTGAC 380
QY 1060 AGCTTGGATGTCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 1119
DB |||||
381 AGCTTGGATGTCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 440
QY 1120 TGTGTACGTGTGCTTTTGGAGAGCTGGTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGT 1179
DB |||||
441 TGTGTACGTGTGCTTTTGGAGAGCTGGTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGT 500
QY 1180 TCTCCCGTGTCTCACAGCTGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1239

Db 501 TCTCCGCTGCTCACAGCTCCCGTGTATGCTGTTGCTATCTCGCAGGAGCTCCTAGAC 560
Qy 1240 CATGTGTCAGAGGCCAACGCTCAAGCTAACTACAGCTGTGGGCTCAAACTAGCTTCA 1299
Db 561 CATGTGTCAGAGGCCAACGCTCAAGCTAACTACAGCTGTGGGCTCAAACTAGCTTCA 620
Qy 1300 TGTTCCTGGCCCTCTATTGTCGCGAGTCTAGGGCAGCTGACCTGTTTCGCGCTGCTT 1359
Db 621 TGTTCCTGGCCCTCTATTGTCGCGAGTCTAGGGCAGCTGACCTGTTTCGCGCTGCTT 680
Qy 1360 TTGCTCCAGGGGCGAGACCTGACTACAACTGCACTGACAGGGCCCTATTGGCTGCTGTC 1419
Db 681 TTGCTCCAGGGGCGAGACCTGACTACAACTGCACTGACAGGGCCCTATTGGCTGCTGTC 740
Qy 1420 CCAAGACCCGCGACCCCTCTTGAATCTGCTCCATCATATAATTTGTGAGCAGATATATC 1479
Db 741 CCAAGACCCGCGACCCCTCTTGAATCTGCTCCATCATATAATTTGTGAGCAGATATATC 800
Qy 1480 CAGCTGTTAATCGATTTTGGTGTCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACC 1539
Db 801 CAGCTGTTAATCGATTTTGGTGTCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACC 860
Qy 1540 TCACAAGATGATAAAGGCAATGTCATGCTGCTACAGGCCCG----- 1580
Db 861 TCACAAGATGATAAAGGCAATGTCATGCTGCTACAGGCCCGAGGTGAGCTGTTCTTCTT 920
Qy 1581 -----AGCACCTCAGGTCATCTTATACAGGTCGGTTAGTCTCGCGAGAGCTTG 1635
Db 921 GCTGTAGCCACTCCAGGTCACCTTCTATCACAGGTCGGTTAGTCTCGCGAGAGCTTG 980
Qy 1636 TGCAGGCTGGCCAGCCACAGCCATCAACAGCTGATATTCCTCCCATGTTGATTAGC 1695
Db 981 TGCAGGCTGGCCAGCCACAGCCATCAACAGCTGATATTCCTCCCATGTTGATTAGC 1040
Qy 1696 TACCTAAAAACCAACTGTAATCTTGACG 1724
Db 1041 TACCTAAAAACCAACTGTAATCTTGACG 1069

RESULT 4
AF403030 927 bp mRNA linear PRI 02-JAN-2002
LOCUS Homo sapiens ankryrin repeat domain-containing SOCS box protein
DEFINITION Asb-12 mRNA, complete cds.
ACCESSION AF403030
VERSION AF403030.1 GI:18034079
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 927)
AUTHORS Kile, B.T. and Nicola, N.A.
TITLE SOCS box proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 927)
AUTHORS Kile, B.T. and Nicola, N.A.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia
FEATURES
source location/Qualifiers
1..927 /organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..927 /codon_start=1
/product="ankryrin repeat domain-containing SOCS box protein Asb-12"
/protein_id="AAL57349.1"
/db_xref="GI:18034080"

/translation="WNLMIDITKIFSLQPKDEEDTDEKQALNQAVYDNDSTLTDQ
LLRBERYKRFINSRSGWPGTPLRLAASVYGHLSCLQVLGAHAGDLSLDVKAQPLF
TAVSHGHLDCVRVLLLEAGSPGGSTVNNCSPLVTAARDGAVAILDELHDHGAENKKA
KLPVWASNIASCSGPLYLAAYVGHLDCEFRLLLLHGDADPDYNTDGLLARVPRPTLL
EICLHNCEPBYIQLLIDFGANITPLPSLSDUTSQDDKGIALLQARATPRSLLSQVR
LWVRALCQAGQPGQAINQLDIPMLISYLKHQ"
BASE COUNT 203 a 274 c 216 g 234 t
ORIGIN
Query Match 49.2%; Score 924; DB 9; Length 927;
Best Local Similarity 100.0%; Pred. No. 1.2e-252;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 787 ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCTGCGCCGACGAGGAGGAG 846
Db 1 ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCTGCGCCGACGAGGAGGAG 60
Qy 847 GACACTGACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 906
Db 61 GACACTGACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
Qy 907 ACTTTGGACAGCTTTTGGCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 966
Db 121 ACTTTGGACAGCTTTTGGCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
Qy 967 TGGGGTGTCTTGGGACACCCCTTGGCTGCTGCTTATGGCCACTTGGAGCTGTTTG 1026
Db 181 TGGGGTGTCTTGGGACACCCCTTGGCTGCTGCTTATGGCCACTTGGAGCTGTTTG 240
Qy 1027 CAAAGTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGAGC 1086
Db 241 CAAAGTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGAGC 300
Qy 1087 CTTTTCACGTGTGTCAGTCATGGCCCATCTGACGTGTGACGTGTGTTGGAAGCTGT 1146
Db 301 CTTTTCACGTGTGTCAGTCATGGCCCATCTGACGTGTGACGTGTGTTGGAAGCTGT 360
Qy 1147 GCCTCTCTGTTGGTAGCATCTACAACTGCTTCCCGTGTCTCACAGCTGCCCGTGT 1206
Db 361 GCCTCTCTGTTGGTAGCATCTACAACTGCTTCCCGTGTCTCACAGCTGCCCGTGT 420
Qy 1207 GGTGCTGTTGCTATCTCTGAGGAGCTCTTAGACCATGTGTGACAGGCCAACGCTCAA 1266
Db 421 GGTGCTGTTGCTATCTCTGAGGAGCTCTTAGACCATGTGTGACAGGCCAACGCTCAA 480
Qy 1267 AAACCTACAGCTGTGGGCATCAAACTAGCTTCTGTCGCTGCTTATTTGGCCGCA 1326
Db 481 AAACCTACAGCTGTGGGCATCAAACTAGCTTCTGTCGCTGCTTATTTGGCCGCA 540
Qy 1327 GTCTACGGGCACTGGAGCTGTTTCCGCTGCTTTCCTCCAGGGGCGAGCCCTGACTAC 1386
Db 541 GTCTACGGGCACTGGAGCTGTTTCCGCTGCTTTCCTCCAGGGGCGAGCCCTGACTAC 600
Qy 1387 AACTGCACTGACAGGGGCTATTGGCTGCTGCTCCCAAGACCCCGACCCCTCTTGA 1446
Db 601 AACTGCACTGACAGGGGCTATTGGCTGCTGCTCCCAAGACCCCGACCCCTCTTGA 660
Qy 1447 TGCCTCCATCATATAATTTGAGCAGAGTATATCCAGCTGTTAATCGATTTTGGTGT 1506
Db 661 TGCCTCCATCATATAATTTGAGCAGAGTATATCCAGCTGTTAATCGATTTTGGTGT 720
Qy 1507 ATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAGGCAATTCAT 1566
Db 721 ATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAGGCAATTCAT 780
Qy 1567 CTGCTACAGGGCCGAGGCCACTCCAGGTCACCTTCTATCACAGGTCGTTTGTCTG 1626
Db 781 CTGCTACAGGGCCGAGGCCACTCCAGGTCACCTTCTATCACAGGTCGTTTGTCTG 840
Qy 1627 AGAGCCTTGTGTCAGGCTGGCCAGCCACAGCCATCAACAGCTGATATTCCTCCCAT 1686
Db 841 AGAGCCTTGTGTCAGGCTGGCCAGCCACAGCCATCAACAGCTGATATTCCTCCCAT 900

```

QY 1687 TTGATTAGTACCTAAACACCAA 1710
|||||
Db 901 TTGATTAGTACCTAAACACCAA 924

RESULT 5
AL356317/c
LOCUS
DEFINITION
  AL356317 104534 bp DNA linear PRI 02-FEB-2001
  Human DNA sequence from clone RP11-284B18 on chromosome Xq12-13.1,
  complete sequence.
ACCESSION
  AL356317
VERSION
  AL356317.8 GI:12666252
KEYWORDS
  HTG.
SOURCE
  Homo sapiens (human)
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 104534)
  Direct Submission
  Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CE10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
  On Feb 5, 2001 this sequence version replaced gi:12578192.
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission
  corresponding to the overlapping clone, as we submit sequences with
  only a small overlap as described above.
  This sequence has been finished according to sequence map criteria
  as follows. An attempt is made to resolve all sequencing problems,
  such as compressions and repeats, but not necessarily within known
  annotated repeat sequence elements. Where the sequence is
  ambiguous, there is an annotation using the 'unsure' feature key.
  The following abbreviations are used to associate primary accession
  numbers given in the feature table with their source databases:
  Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
  on the WORMPEP database can be found at
  http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
  was generated from part of bacterial clone contigs of human
  chromosome X, constructed by the Sanger Centre Chromosome X Mapping
  Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/ChrX
  RP11-284B18 is from the library RPC1-11.1 constructed by the group
  of Pieter de Jong. For further details see
  http://www.chori.org/bacpac/home.htm
  VECTOR: pBACe3.6
  IMPORTANT: This sequence is not the entire insert of clone
  RP11-284B18 It may be shorter because we sequence overlapping
  sections only once, except for a 100 base overlap.
  The true left end of clone RP11-284B18 is at 1 in this sequence.
  The true left end of clone RP4-710L4 is at 104435 in this sequence.
  The true right end of clone RP11-403E24 is at 63843 in this
  sequence.

FEATURES
    source
    1..104534
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosomes="X"
    /map="q12-13.1"
    /clone="RP11-284B18"
    /clone_lib="RPC1-11.1"
    1..4482
    /notes="HERVL repeat: matches 1..4542 of consensus"
    4489..5037
    /note="MLT2B repeat: matches 1..438 of consensus"
    5038..5133
    /notes="LIPAL3 repeat: matches 5960..6047 of consensus"
    5139..5388
    /note="LIPAL3 repeat: matches 480..5401 of consensus"
    5389..6274
    /note="L2 repeat: matches 2689..2746 of consensus"
    6136..6274

    repeat_region
    /note="MIR repeat: matches 49..203 of consensus"
    6859..7220
    /note="MLT1A1 repeat: matches 1..365 of consensus"
    8549..8744
    /note="LTR33 repeat: matches 6..216 of consensus"
    12540..12716
    /note="MIR repeat: matches 20..201 of consensus"
    12732..13094
    /note="L2 repeat: matches 1129..1502 of consensus"
    13311..13725
    /note="L2 repeat: matches 1864..2295 of consensus"
    13964..14156
    /note="MIR repeat: matches 58..261 of consensus"
    14257..14330
    /note="MIR repeat: matches 79..152 of consensus"
    14386..14524
    /note="MIR repeat: matches 47..183 of consensus"
    15029..15467
    /note="Charlie4a repeat: matches 25..507 of consensus"
    16483..17040
    /note="L1P4 repeat: matches 5585..6142 of consensus"
    17775..18439
    /note="L1M2 repeat: matches 5677..6341 of consensus"
    18432..18880
    /note="L1M4 repeat: matches 4317..4766 of consensus"
    18881..19351
    /note="L1P8 repeat: matches 5682..6163 of consensus"
    19352..19362
    /note="L1M4 repeat: matches 4307..4317 of consensus"
    19465..20389
    /note="L1M4 repeat: matches 3330..4263 of consensus"
    20397..20452
    /note="28 copies 2 mer ta 87% conserved"
    20481..21323
    /note="L1M5 repeat: matches 5313..6138 of consensus"
    21799..22116
    /note="L1M2 repeat: matches 5844..6160 of consensus"
    22130..22456
    /note="MSTA repeat: matches 55..426 of consensus"
    22457..22511
    /note="MLT2B repeat: matches 390..444 of consensus"
    22512..22535
    /note="12 copies 2 mer ta 100% conserved"
    22615..22989
    /note="MLT2B repeat: matches 1..382 of consensus"
    22990..23047
    /note="MSTA repeat: matches 1..60 of consensus"
    23048..23953
    /note="L1M6 repeat: matches 4929..5842 of consensus"
    23954..24953
    /note="MER11C repeat: matches 1..1071 of consensus"
    24955..25128
    /note="L1M6 repeat: matches 4759..4933 of consensus"
    25155..25280
    /note="L1 repeat: matches 4020..4145 of consensus"
    25280..25431
    /note="L1 repeat: matches 5175..5334 of consensus"
    25592..26020
    /note="L1 repeat: matches 4767..5200 of consensus"
    26031..26260
    /note="L1M4A repeat: matches 5987..6274 of consensus"
    26261..26667
    /note="MSTB repeat: matches 2..422 of consensus"
    26673..27291
    /note="L1M4A repeat: matches 5318..5988 of consensus"
    27284..28932
    /note="L1 repeat: matches 3281..4957 of consensus"
    28933..28976
    /note="L1P13 repeat: matches 6113..6156 of consensus"
    28985..33897
    /note="L1P13 repeat: matches 480..5401 of consensus"
    34102..34982
    /note="L1P13 repeat: matches -651..215 of consensus"

```


JOURNAL	Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia
FEATURES	Location/Qualifiers
source	1. .1262
	/organism="Mus musculus"
	/mol_type="mRNA"
	/db_xref="taxon:10090"
CDS	198..1124
	/codon_start=1
	/product="ankyrin repeat domain-containing SOCS box protein Abb-12"
	/protein_id="AA157359.1"
	/db_xref="GI:18034100"
	/translation="NNLMDIAKIFSLLOPEKEEDTGTGCKEALNQAQVVDNSCLDLHLLHQRYKRFINSRSGWGPPTPLRAASYGHLNCKVLLHGHADVSLDVKQAQPLF TAVSHGUELVNLEAGACPSGSIYNNCSFVLTASRDGAFALBELGLHGHAEANVKA KLPVWASNIASCGLYLAAYVGHLDPCFLLLLYAGADPDYNCITDGLLSRVPPQRTLL EICLHNCEPEYIQLIDFGANIYLPSPVDPSTQDDKGIKLLQARATPRSLSSQTR LVIRSLCRANQSQATDQIDIPVLISYLKHQ"
BASE COUNT	318 a 338 c 272 g 334 t
ORIGIN	
	Query Match 38.8%; Score 729.2; DB 10; Length 1262;
	Best Local Similarity 81.5%; Pred. No. 6.9e-197;
	Matches 883; Conservative 0; Mismatches 193; Indels 8; Gaps 3;
QY	736 ATCTGCAATTCCTTGTGCTCATATAG--AATAGTTCTCCAAATTAGCCAAGATGAACC 793
DB	145 ATCTGAAATTTCTCTCTCTTGGCAATGAGATGCTATTTCGAATAAACCAAGAAATGAAC 204
QY	794 TCATGACATCACCAAGATCTTCTCCCTCTCTGAGCCGACAGAGGAGGAGGACACTG 853
DB	205 TCATGATATGCCAAGATCTTCTCTCTGCACTGAAAGAGGAGGAGGACACTG 264
QY	854 ACACAGAGGAGAGCGGCTCTCAATCAAGCAGTGTATGACACAGCTCTCTATCTTTGG 913
DB	265 GCACCGGGGAAACAGGCTCTCAATCAAGCTGTATGACATGACTCTGTACCTGG 324
QY	914 ACCAGTTTGGCCAGGAGCGTTACAAAGCTTTATCAACAGCAGGAGTGGCTGGGGTG 973
DB	325 ACCACCTTACACAGGACGTTATAACGGTTTCATCAACAGCAGGAGTGGCTGGGGTA 384
QY	974 TTCCTGGGACACCTTGGCTTGGCTTCTTATGCGCACTTGAGCTGTGTTGCAAGTCC 1033
DB	385 TACCTGGAACACCTTGGCTTGGCAGCTTCTTATGCTCACTTAAATTTGTGAAGGCTC 444
QY	1034 TCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAGAGCCCACTTTTCA 1093
DB	445 TCTTGAACATGTGTGATGTTGATGCTTGGATGTCAAGGCACACACCACTTTTCA 504
QY	1094 CTGCTGTAGTCAATGCCATCTGGACTGTGTAGTGTGCTGTTTGGAGCTGGTCTCTC 1153
DB	505 CTGCTGTAGCCACCGTCTCTGGAGTGTGAGAAATGCTTTTGAAGCTGGTGGCTGTC 564
QY	1154 CTGGTGTAGCATCTACAACTGTTTCCCGTGTCTACAGCTGCCCGTGTGTTGCTG 1213
DB	565 CTAGTGTAGCATCTACAACTGTTTCCCGTGTCTACAGCTGCCCGTGTGTTGCTG 624
QY	1214 TTCTATCTCTGAGGAGCTCTAGACCATGGTGTGAGAGGCCAACGTCAAAGCTTAAACTAC 1273
DB	625 TTGCCATCTTACAGGAGCTCTAGGCAATGTTCCCGAGGCTATGTCAAGCTTAACTAC 684
QY	1274 CAGTCTGGGCATCAAAATAGCTTTCATGTTCTGGCCCTCTATTTTGGCCGAGTCTACG 1333
DB	685 CAGTCTGGGCGTCAAAATATAGCTTTCATGTTCTGGCCCTCTATTTTGGCTGAGTCTATG 744
QY	1334 GGCACCTGAGCTTTTCCGCTCTTTTGTCTCCACGGGCGAGACCTGCTACTACACTGCA 1393
DB	745 GGCACCTTGAATTTTCCGCTCTTTTGTCTATGGGGCAGATCTCTGATTACACTGCA 804
QY	1394 CTGACACGGGCTTATTTGGCTGTGTCCTCAAGACCCCGCAGCTCTCTTGAATCTGCTCC 1453
Db	805 CTGACACGGGCTTTTAAAGTCGTGTTCACAGCCTCCGACACTCTCTTGAATCTGCTTC 864
QY	1454 ATCATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTCGTGCTAATATCTACC 1513
Db	865 ATCATAATTTGTGAGCCAGAGTATATCCAGCTGTTTAAATAGATTTCGTGAGCTAATCTACC 924
QY	1514 TTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGGCATTTGATGCTGTAC 1573
Db	925 TTCCATCTCTCTCTGTGGACCCAACTTCACAAGATGATAAAGGCATCAAAATGCTGTAC 984
QY	1574 AGCCCGAGGACCACTCCAGCTCACTTCTATCACAGTCCGTTAGTCGTCGAGAGCCT 1633
Db	985 AAGCCCGAGGACCACTCCAGCTCACTTCTGTCCAGACCCGTTTGTATTCGCGAGATCC 1044
QY	1634 TGTGCGAGGCTGGCCAGCCACCAAGCATCAACAGCTGGATATTCCTCCCATGTTGATTA 1693
Db	1045 TCTGCGGGGACCAAGTCAACAAGCCAGGACCTGGATATCCCTCTGTGTTGATTA 1104
QY	1694 GTTACCTAAACACCAACTGTAATCTTCAGTCTCCCCAGGAACCTT-----ATGATGCCT 1748
Db	1105 GTTACCTCAACATCAATGATAAGTTTCACTTTCTCAGGAACCTAATGCTATGATACCT 1164
QY	1749 CCGAAAA--CCACTGGGACTCACTAGCTGGAGCATTACAGCTCATCCACTTACCT 1807
Db	1165 CCAAAAAATAAACCTGGTAACCCAGGTTACTTGAGTGAATAATACACCTCATAACTTCT 1224
QY	1808 GGAG 1811
Db	1225 GGCG 1228
RESULT 7	
BC012399	Homo sapiens, Similar to hypothetical protein FLJ20126, clone
LOCUS	MGC:8876 IMAGE:3864128, mRNA, complete cds.
DEFINITION	
ACCESSION	BC012399
VERSION	BC012399.1 GI:15214550
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 2653)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (15-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgabbs@mail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villalon@bcm.tmc.edu Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.
FEATURES	
source	1. .2653
	/organism="Homo sapiens"
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 21 Row: j Column: 23 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7020020. Location/Qualifiers 1. .2653

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:8876 IMAGE:3864128"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_66"
/lab_host="DH10B"
/notes="vector: pcwv-sport6"
69..2183
CDS
    /codon_start=1
    /product="Similar to hypothetical protein FLJ20126"
    /protein_id="AAH12399.1"
    /db_xref="GI:15214551"
    /translation="MDHITVPKVENKLVDRYVSKPANGILYLTATHLIYVEASGAA
    RKETIALHIAIVEKLPITSLGCLPTLRCKNFRVAHFVLDSDLVCHVEYIISLLKLSQ
    PALPDLVAFSPYKSKEMRSKWKLDIPISDFGRMGI PNRNWITTDARNVEICST
    YPEEIVVPSVTLGTGVGSKFRKERVPLVLYKNNAAICRCQPSLGSFVTRCVD
    DELLLEAISOTNPGSQFMYVVDTRPKLNAMANAAGKGVENEDNYAIRFPWGIENI
    HVMRSIOLKLEVECEKLTPTMSFSLGSSGWRHKAIMDAGIFITKAVKVEKASV
    LVHCSQDMORTAQVCSVASIILLDPFRTFKGLMLIEKEMWSMGRKFSQRCGLDGS
    KEVSPITQDLDCIQWMEQPCAFNFENFLEIHDHVFPSQCFQNGFLNCQKQREDL
    RVYEKTHSVWPFVQRKPPFRNPLYKFTMYGVLPSTVYPIQFQWCMYRNFGLQ
    PKOSMLELLEIKQAMLETVDHELEKLVKRDPEPESICTCSQGLNLSLHGLSPL
    TNPGLFWMGINDLNTMENGTLREGGLRAQMDQVKSOGADLHHNCEIVGSLRAINI
    SGDVGISSEAMGISDGMTPEATGFSKDLGICGAMDISEATGISGNLIGISEARFSDGM
    GILGDTGISKATKEADYSKHQ"
BASE COUNT 749 a 566 c 645 g 693 t
ORIGIN
    Query Match 37.8%; Score 710.2; DB 9; Length 2653;
    Best Local Similarity 99.6%; Pred. No. 2e-191, 3; Indels 0; Gaps 0;
    Matches 712; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 23 CGAGGCTAGAAAACGTAATGGTGGATCGTTATGTGAGTAAGAAACACAGCTAATGGGA 82
DB 88 CCAAGGTAGAAAACGTAATGGTGGATCGTTATGTGAGTAAGAAACACAGCTAATGGGA 147
QY 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 142
DB 148 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 207
QY 143 AAACATGATGATGACCTCCATCAATGCCATCTGAGAGAGTTACCCATCAGTACGCTGG 202
DB 208 AAACATGATGATGACCTCCATCAATGCCATCTGAGAGAGTTACCCATCAGTACGCTGG 267
QY 203 GTTGTCCTCCAGCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 262
DB 268 GTTGTCCTCCAGCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 327
QY 263 ACCTTGTCGCATGAGGTTTATATTTCTACTGCTCAAGCTTTCTAGCAGCAATACCTG 322
DB 328 ACCTTGTCGCATGAGGTTTATATTTCTACTGCTCAAGCTTTCTAGCAGCAATACCTG 387
QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGAAAGTGAT 382
DB 388 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGAAAGTGAT 447
QY 383 GGAACCTGATGACCCCAATATCAGACTTTGGGCGTATGGGAATFACCCACAGAAACTGGA 442
DB 448 GGAACCTGATGACCCCAATATCAGACTTTGGGCGTATGGGAATFACCCACAGAAACTGGA 507
QY 443 CCATAACAGATGCCAACAGAACTATGAGATATGAGACCTTACCTCTCGTAAATAGTGG 502
DB 508 CCATAACAGATGCCAACAGAACTATGAGATATGAGACCTTACCTCTCGTAAATAGTGG 567
QY 503 TTCTCTAAATCTGTTTACCTTGGGAACGGTGTGGAAAGTTCAAAGTTCAAGATGAAGAAC 562
DB 568 TTCTCTAAATCTGTTTACCTTGGGAACGGTGTGGAAAGTTCAAAGTTCAAGATGAAGAAC 627
QY 563 GTGTCCCTGTGCTCTCTACCTCTACAAAGAGAAACAAATGCTGCCATTTGCCCGCTGTAGCC 622
DB 628 GTGTCCCTGTGCTCTCTACCTCTACAAAGAGAAACAAATGCTGCCATTTGCCCGCTGTAGCC 687
QY 623 AGCCTCTCTCGGATTTTACATCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTGA 682
```

```
|||||
688 AGCCTCTCTCTGATTTTAACTCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTA 747
|||||
683 GCCAACAACACCCAGGAGCCAGTTTATGTATCTTGTAGACACAGACCAAGAT 737
|||||
748 GCCAACAACACCCAGGAGCCAGTTTATGTATCTTGTAGACACAGACCAAGAT 802
|||||

RESULT 8
AX406008 1345 bp DNA linear PAT 14-JUN-2002
LOCUS AX406008
DEFINITION Sequence 423 from Patent WO0222660.
ACCESSION AX406008
VERSION AX406008.1 GI:21439455
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Mehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 423 21-MAR-2002;
HYSBO, INC. (US)
FEATURES
    Location/Qualifiers
    1..1345
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    92..1204
    /note="unnamed protein product"
    /codon_start=1
    /protein_id="CAD35115.1"
    /db_xref="GI:21439456"
    /translation="MDHITVPKVENKLVDRYVSKPANGILYLTATHLIYVEASGAA
    RKETIALHIAIVEKLPITSLGCLPTLRCKNFRVAHFVLDSDLVCHVEYIISLLKLSQ
    PALPDLVAFSPYKSKEMRSKWKLDIPISDFGRMGI PNRNWITTDARNVEICST
    YPEEIVVPSVTLGTGVGSKFRKERVPLVLYKNNAAICRCQPSLGSFVTRCVD
    DELLLEAISOTNPGSQFMYVVDTRPKAVKVEKASVLVHCSQDMORTAQVCSVASIILLD
    PPFRTFKGLMLIEKEMWSMGRKFSQRCGLDGSKEVSPITQDLDCIQWMEQPCAFNFENFLEI
    HDHVFPSQCFQNGFLNCQKQREDLR"
BASE COUNT 380 a 300 c 321 g 344 t
ORIGIN
    Query Match 37.5%; Score 703.4; DB 6; Length 1345;
    Best Local Similarity 99.2%; Pred. No. 1.7e-189;
    Matches 707; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 23 CGAGGCTAGAAAACGTAATGGTGGATCGTTATGTGAGTAAGAAACACAGCTAATGGGA 82
DB 111 CCAAGGTAGAAAACGTAATGGTGGATCGTTATGTGAGTAAGAAACACAGCTAATGGGA 170
QY 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 142
DB 171 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 230
QY 143 AAACATGATGATGACCTCCATCAATGCCATCTGAGAGAGTTACCCATCAGTACGCTGG 202
DB 231 AAACATGATGATGACCTCCATCAATGCCATCTGAGAGAGTTACCCATCAGTACGCTGG 290
QY 203 GTTGTCCTCCAGCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 262
DB 291 GTTGTCCTCCAGCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 350
QY 263 ACCTTGTCGCATGAGGTTTATATTTCTACTGCTCAAGCTTTCTAGCAGCAATACCTG 322
DB 351 ACCTTGTCGCATGAGGTTTATATTTCTACTGCTCAAGCTTTCTAGCAGCAATACCTG 410
QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGAAAGTGAT 382
DB 411 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGAAAGTGAT 470
```



```

RESULT 10
AC109692
LOCUS
DEFINITION
AC109692
VERSION
AC109692.4
KEYWORDS
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 194079)
AUTHORS
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Altschuler,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Mileasavjevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savory,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soza,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,L., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldfon,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczkyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
REFERENCE
2 (bases 1 to 194079)
AUTHORS
Worley,K.C.
JOURNAL
Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
3 (bases 1 to 194079)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21737980.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQBH
Center clone name: CH230-268J15
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 167200 bases at least Q40
Consensus quality: 170449 bases at least Q30
Consensus quality: 172661 bases at least Q20
Estimated insert size: 186192; sum-of-contigs estimation
Quality coverage: 4x in Q20.bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194079: contig of 194079 bp in length.
*
FEATURES
source
1. 194079
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-268J15"
1. 1257
/note="wgs end_extension
clone end:77"
6621..7521
/note="clone_boundary
clone end:77
site:MboI
end sequence:RXAE056TV"
110101..112444
/note="wgs_contig"
192195..193062
/note="clone_boundary
clone end:Spf
site:MboI
end sequence:RXAE056TV"
BASE COUNT 48623 a 34416 c 35611 g 55333 t 20096 others
ORIGIN

Query Match 32.6%; Score 612; DB 2; Length 194079;
Best Local Similarity 83.2%; Pred. No. 4.1e-163;

```

Matches 709; Conservative 0; Mismatches 140; Indels 3; Gaps 1;	
QY 734	AGATCTGGCATTTCTTGTGCTCATATAGAGATAGTTC---TCCAAATAGCAAGATCA 790
Db 152248	AGATCGGGAATTTCTTCTCTTTTACATAGAGATAGCTCTATTCNATTAACAGATGA 152307
QY 791	ACCTCATGACATCAACCAAGATCTTCTCCCTCTGAGCCGACCAAGGAGGAGGACA 850
Db 152308	ACCTCATGATATCGCAAGATCTTCTCTCTTCTGCAACCAAGGAGGAGGAGGAG 152367
QY 851	CTGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATCAACAGCTCTATATCT 910
Db 152368	CCGACACAGGCGAAAGAGGCTCTCAATCAAGCTGTATATGACATGACTCTTGTACC 152427
QY 911	TGACACAGCTTTTGGCCAGGAGCGTTACAAAGTTTCAACACAGCAGGATGGCTGG 970
Db 152428	TGACACACCTTTACACAGGAGGTTATTAACGGTTTCAACACAGCAGGATGGCTGG 152487
QY 971	GTGTTCTCTGGACACCTTTGGCTTGGCTGCTTCTTATGGCCACTTGAAGCTGTG 1030
Db 152488	GTATACCTTGGACACCTTTGGCTTGGCAGCTTATGTCACCTTAGATTGTGGAAG 152547
QY 1031	TCCTCTTAGCCATGTGCTGATGTGACAGCTTGGATGTCAAGGACAGACGCACTTT 1090
Db 152548	TCCTCTGGAACATGTGCTGATGTGATGATGCTTGAAGCTTGAAGCACAACACCACTTT 152607
QY 1091	TCACTGCTGTGATGATGCGCATCTGCACTGTGTAGTGTCTTTTGGAGCTGTGGCT 1150
Db 152608	TCACTGCTGTGTACGGTCACTGTGAGTGTGTGAGATCTTTTGAAGCTGTGGCT 152667
QY 1151	CTCTGCTGTGATGATCTCAACAACTGTTCTCCCTGCTGCAGCTGCCCTGTGATGG 1210
Db 152668	GTCTGATGTGATGATCTCAACAACTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 152727
QY 1211	CTGTTGCTATCTGAGAGCTCTTAGACATGCTGTCAGAGGCAACGCTCAAGCTAAAC 1270
Db 152728	CCTTTGCCATCTTACAGGAGCTCTTAGGCGATGCTGAGAGGCTATGTCAAGGCGAAG 152787
QY 1271	TACCACTGCGGCATCAACATAGCTTCATGTTCTGCGCCCTCTATTTGGCGAGTCT 1330
Db 152788	TGCACTGTGGGATCAACATAGCTTCATGTTCTGCGCCCTCTATTTGGCGAGTCT 152847
QY 1331	AGGGCAGCTGCACTGTTTCCGCTCTCTTTTGTCTCAACGGGGCAGACCTGATCAACT 1390
Db 152848	ATGGTATCTTGTATGTTTTCGCTCTCTTGTCTATGGGCGAGATCTGATTACACT 152907
QY 1391	GCATGACAGGCGCTTATGGCTGTGTCGCAAGACCCCGCACCTCTCTGAAATCTGCC 1450
Db 152908	GTATTGACAGGCGCTTAAAGTCGAGTTCACAGGCTCGCACACTCTCTGAAATCTGCC 152967
QY 1451	TCATCATATTTGTGAGCCAGATATATCCAGCTGTTATCGATTTTGTGCTATATCT 1510
Db 152968	TTATCATATTTGTGAGCCAGATATATATAGCTTTTATAGATTTTGGAGCGAATCT 153027
QY 1511	ACCTTCCATCT 1570
Db 153028	ACCTTCCATCT 153087
QY 1571	TACAGGCGCGAG 1582
Db 153088	TACAAGCGCGAG 153099

RESULT 11

AC108644	
LOCUS	Rattus norvegicus clone CH230-59A6, *** SEQUENCING IN PROGRESS ***
DEFINITION	8 unordered pieces.
ACCESSION	AC108644
VERSION	AC108644.5 GI:24818595
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 225982)

Muzny,D., Marie, Metker, M., Lee, A., Abruzzo, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., De Souza, L., Devila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guetara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshwar, L., Louisedge, H., Lozano, R. J., Lu, X., Ma, J., Maheshwar, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okwionu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallas, R., Vers, V., Villalana, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 240816)
Sehra, H.
Direct Submission
Submitted (11-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2003 this sequence version replaced gi:27652710.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfsh-help@sanger.ac.uk

COMMENT

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep Repeat names
beginning 'Dr' were identified by the Recon repeat discovery system
(Zhiron Bao and Sean Eddy, submitted), and those beginning 'dr'
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
further information see http://www/Projects/D_rerio/fishmask.shtml
DKEY-20D18 is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.

FEATURES

source

1..240816
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-20D18"
/clone_lib="DanioKey"
/clone="DKEY-20D18"

BASE COUNT 76749 a 43368 c 43554 g 77145 t
ORIGIN

Query Match 15.5%; Score 290.6; DB 5; Length 240816;
Best Local Similarity 61.0%; Pred. No. 3.9e-71;
Matches 489; Conservative 0; Mismatches 309; Indels 3; Gaps 1;
QY 782 CCAAGATGAACCTCATGGACATCAACAAGATCTTCTCCTCGACGCGCAAGAGG 941
DB 18732 CCAGCATGATGTTGATGGACATCTCTAAGATCTTTCTCTTCAACCAAGAGGATG 18791
QY 842 AGGAGGACACTGCACAGGAGGAGGAGGCTCTCAATCAAGCAGTGTATGACAGCT 901
DB 18792 AGGAGGAA---GACACCAACTGCAGTCAAGCACTAATCAAGTGTCTTCAAGAGCAG 18848
QY 902 CTTATCTTTGGACCAAGCTTTTTCGCGCAGGAGGTTTCAAAAGCTTTCATCAACAGCAGGA 961
DB 18849 ACAAACTCTTACTGAATGTTGTCTCAGGACAAATATAGAAAGTGCATCAACACGCA 18908
QY 962 GTGGCTGGGGTGTCTTGGGACACCTTGGCTGTGCTCTTATGACCACTTGAAGCT 1021
DB 18909 GTGGTGGGGGATCCCACTCACTCCACTGCGCAGCGGGCTGCTCAGGGTCACTACGCT 18968
QY 1022 GTTGAAGTCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAGGACAGCA 1081
DB 18969 GTCTGAGGTCTCTCTGGCTCAAGGGGAGAGGTCGACGCTGGATGTCAAAGCTCAAA 19028
QY 1082 CGCCACTTTTCACTGTCTGCTCAGTCATGGCCATCTGGAATGTGTACGTGTCTTTTGAAG 1141

Db 19029 CACCACCTTTTACAGCCGCTCTCTGCAAAACACATAGATTGTGTGCTTTATTAAAGAG 19088
QY 1142 CTGGTGCTCTCTCGTGGTAGCATCTACAACTGTCTCCCGTGTCTCAGCTGCCCC 1201
Db 19089 CTGGGCGAGACCTTAATGAAGCCCAACATAAACAATGCTCCCTGTGTGTGACTGGGCA 19148
QY 1202 GTGATGGTGTCTTCTATCTCTGAGGAGCTCTTAGACCATGCTGTCAGAGGCAAGCTCA 1261
Db 19149 GAGAGGGGATGTGCACATCTCTGAAGAGCTCTTCAATATGCGCCGATGAGAGCTCA 19208
QY 1262 AAGCTAAACTACAGCTGCGGCATCAACATAGCTTTCATGTTCTGCGCCCTCTATTGG 1321
Db 19209 AGCCCCAAATGCGACAGCTGGGCTTTAAATGCGCACAGCTTGGCGAGGACCACTGTATATT 19268
QY 1322 CGCAGCTCTAGGGCAGCTGAGCTGCTTCCGCTCTTCTGCTCCAGGGGAGACCCCTG 1381
Db 19269 CAGCGGTGTATGTCATCTAGGTTGTCTTCAATGTTGTCGAGTTTGGAGCAACCCAG 19328
QY 1382 ACTACAACTGCACTGACAGGCGCTATTGGCTGCTGTCCTCAAGACCCCGACCTCTCTTG 1441
Db 19329 ACTACAACTGCACTGAGGAAGATGCTGGCCAGGATCAAGAGGCTTAAACCTGTTTGG 19388
QY 1442 AAATCTGCTCATCAATATGTCAGCCAGATATATCCAGCTGTTAATCGATTTTGGTG 1501
Db 19389 AGATGTGTCTGAGACATGCTGTGGCTGGAATACATTCAGCTGTCTATAGACTTTGGAG 19448
QY 1502 CTAATATCTACCTCTCATCTCTCTCCCTGACCTGACCTCACAAGATGATAAGGATTTG 1561
Db 19449 CAAATGTTTATCTGCCCCCTTTAGTTGGAGATAAACCCAGTGAAGATGAAGCAGTAG 19508
QY 1562 CATTCGTCTGCTACAGGCCCGAG 1582
Db 19509 TCTTCTGCTTAAAGAGAGAG 19529

RESULT 13

BC053411

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

1 (bases 1 to 2354)

AUTHORS
TITLE
JOURNAL

Strausberg, R.
Direct Submission
Submitted (02-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsh-x@mail.nih.gov
Tissue Procurement: Dr. Sumio Sugano
cDNA Library Preparation: Dr. Sumio Sugano
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca

REMARK
COMMENT

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 104 Row: c Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GeneScan gene prediction, Similarity but not identity
to protein.

FEATURES

source

1. .2354

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="MGC:56405 IMAGE:5605218"

/tissue_type="whole body, adult male"

/clone_lib="Sugano SJD adult male"

/lab_host="DH10B"

/notes="Vector: pME18S-FL3"

147. .2045

/codon_start=1

/product="Unknown (protein for MGC:56405)"

/protein_id="AAH53411.1"

/db_xref="GI:31418218"

/translations="MEHIIPTKVENYKLLNRYTEKKSALGTLYLTHALIIYVEQTSNT
RKEAWLHHHILSVKLLTAGSCPLLIRCKTPQHLHLFLQKRDQDVYQSLRLSQ
PKVEELVAFLYNPHQNEERRRGGWELISVNDNFMGLSNDYWEIASHINKPNCST
YPAILGLPKASVATVTGSAKPSRGRPLVLSYHKDKAAICRCQSPGLNSRCVE
DEOMLOAIQANPNSPIYVVDPRPKLNAMNPAAGKGYENEDNYSNIRFQGLIENI
HVRSSLOKLLVCSSKMSRMSDYLTLGLENGLRHLIKSVMDAGVFLAKAVEERASV
LVHSDGWRDTAGVCSLACLLDPFYRTIKGLMVLIEKEWISFGHKFSPHRCGLSDP
KEASPVFTQECVWLSQFPVCFEFNEFHEIYHIDVYACQYGNFIGNYCKERLDM
RLHEKFTLWPHLEHQVRNPLRSLAVSTLVRPSTPLPLHFKPWCNGYNNHYDRGMH
PKQSLVDLTLLTOROVEGERTWELQROLAVADGVLPPAGPINTHGDONNOSKMP
APVQVNSCAPLINGNVKPGVGAENSNQEDREBPANEDHLSKDKPVSVEIHS
KVEVQSS"

BASE COUNT 707 a 519 c 552 g 576 t

ORIGIN

Query Match 13.5%; Score 254.2; DB 5; Length 2354;

Best Local Similarity 59.7%; Pred. No. 6.2e-61;

Matches 427; Conservative 0; Mismatches 288; Indels 0; Gaps 0;

QY 23 CGAGGGTAGAAAACGTGAATTCGGTATCGTATGTGAGTAAAGAACACAGCTAATGGGA 82

Db 166 CGAAGGTGGAGATGTGAACACTACTGAACAGGTACACTGAAAGAGAGTCAAGCTCGGCA 225

QY 83 TTCTTTATCTTACTGCAACCCACTGATCTATGTGGAGGCTTCAGGTGACGCCGGAAG 142

Db 226 CTCTTTACTTGAAGTCACTATCTATCTATGTGGAACAAACAGCAACACGCAAG 285

QY 143 AAACATGGATGGCACTCCATCATCATCTGTCACCTGTGGAGAAGTTACCCATCACTAGCCTGG 202

Db 286 AAGCATGGGTGCTTTCATCATCATCTGTCAGTGGAGAAACTTCTGTTAACTGCGTCCG 345

QY 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTCGGGTGGGCCACATTTGTTTATAGATCTG 262

Db 346 GTTGTCCACTGTTGATCCGCTGTAAACCTTCCAGCATCTGCATCTGCTGTTTCAAAAAG 405

QY 263 ACCTGTGTGCGCATGAGGTTTATTTCACTGCTCAAGCTTTTCTCAGCCAGCATTAACCTG 322

Db 406 AGAGGATTTGTGAGGATGTGATCAGTCACTTCTACGGCTTTTCAAACTGTCAAAGAAG 465

QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAAAATCTCTCAAAAGAGATGAGGGAAGTGGAT 382

Db 466 AGGAGCTTTATGCTTTCTCTATAACCCACATCAAAACGAGGAGGAGCGGCGGTT 525

QY 383 GGAACACTGATGACCCCAATATCAGACTTTGGGGGTATGGGAATACCCACAGAACTGGA 442

Db 526 GGGAGTTGATCAGTGTGGTGAATGACTTCAACAGGATGGGCTCTCTAATGATTAATTGGG 585

QY 443 CCATAACAGATGCCACAGAACTATGAGATATGACAGACCTACCTCCTGAAATAGTGG 502

Db 586 AAATTAGCCATATCAACAAAAACTTTGAGATGTGTAGCACTTATCCAGCATCTCTAGGAC 645

QY 503 TTCTAAATCTGTGTACCTTTGGGAAACGGTGGTGGAAAGTTCAAAGTTTCAGAAAGTAAAGAAC 562

Db 646 TCCTTAAAGTGCAGTGTGTACGGTAAACAGGAGTGCCTAAATTTAGAGCCGAGGGC 705

QY 563 GTGTCCCTGTGCTCTCACTCTCTCAAAAGAGAACTGCTGCAATTTGCGGCTGTAGCC 622

Db 706 GTTGGCTGTTCTGTCTACTATCAAAAGATACAAAGGCTGCGATTTGTGCTGCAATC 765

QY 623 AGCTCTCTCTGATTTTACACTCGCTGTGTAGATGATGAGCTCTGTTGGAGGCCATTA 682

Db 766 AGCTTTGTTCAGGTTTGAATCCCGCTGTGTGGAGACGAGGAGATGCTTCAGGCCATCA 825

QY 683 GCCAAACAAACCCAGGAGCCAGTGTATGTATGTTGTAGACACAAACCAAGAT 737

Db 826 GCCAGCCAAATCCAAACTCTCCCTTTATATATGTAGTGACACACACGCGCCCAAGTT 880

RESULT 14

AX086038

LOCUS

Sequence 35 from Patent WO0112819.

AX086038

ACCESSION

AX086038.1 GI:13275850

VERSION

KEYWORDS

Source

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

Plowman, G.D., Martinez, R., Whyte, D., Hill, R., Flanagan, P. and

Lioubin, M.

Protein phosphatases and diagnosis and treatment of

phosphatase-related disorders

Patent: WO 0112819-A 35 22-FEB-2001;

Sugen, Inc. (US)

Location/Qualifiers

1. .2353

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 670 a 506 c 535 g 642 t

ORIGIN

Query Match 12.9%; Score 241.8; DB 6; Length 2353;

Best Local Similarity 58.3%; Pred. No. 2.2e-57;

Matches 423; Conservative 0; Mismatches 302; Indels 0; Gaps 0;

QY 10 GCCAAGCTTGGCAGCGGCTAGAAAACGTGAAATTTGGTGGATCGTTATGTGAGTAAAGAA 69

```
Db 239 GCTCAGGGAAGCGGAAAGGTTGAAATGTCGCGTTGGTAGATCGAGTGTCTCTCTAAAAA 298
QY 70 CCAGCTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGT 129
Db 299 GCAGCTCTAGGTACTTTGTTATTTGACGGCTACCCATGTCTATATGTTGGAAATTCACCT 358
QY 130 GCAGCCCGGAAAGAAACATGGATTGCACTCCATCACAATTGCCACTGTGGAGAAGTTTACCC 189
Db 359 GAGCGAAGAAAGAAACATGGATTCTTCACAGTCAGATTTCCACCAATTGAGAAACAGGCA 418
QY 190 ATCACTAGCTGGTGTGCTCCCTGACCCCTCCGTCGAAGAAATTCGGGTGGCCCACTTT 249
Db 419 ACAACCGCTACCGGATGCCCTCTGCTGATTCGTCGAAGAACTTTTCAGATAATACAGCTC 478
QY 250 GTTTTAGATTCTGACCTGTGTGCGCATGAGTTTATATTTCATCTGCTCAAGCTTTCTCAG 309
Db 479 ATCATACCTCAGGAAGAGATTGCCACGAGTGTATCTCTCCCTGATACGCTTGAAGG 538
QY 310 CCAGCATTTACCTGAAGATCTTTATGCTTTTCTTATAATCCCAATCCCTCAAAAGAGATG 369
Db 539 CCAGTGAATATGAGGAGTTATACTGCTTTTCATTCAACCCCATGCTGGATAAAGAGAA 598
QY 370 AGGAAAGTGGATGGAACATGATTGACCAATATCAGACTTTTGGGGGTATGGGAATACCC 429
Db 599 AGAGCAAGGCTGGGTGCTGATCGATCTTAGTGAAGAAATACACGGGATGGGCTCCCT 658
QY 430 AACAGAACTGGACATAACAGATGCCACAGAACTATGAGATATGACGACCTACCT 489
Db 659 AATCATTTACTGGCAGCTCAGCGATGTAATAGAGACTATAGAGTCTGTGACTCTTATCT 718
QY 490 CTTGAATAGTGGTCTTAAATCTGTACTCTTGGGAACGGTGTGGAAAGTTCAAAGTTC 549
Db 719 ACTGAATCTGACTTCCCAATCGGCCACGACACATATAGTGGGAGTTCCAAATTC 778
QY 550 AGAAGTAAAGAACGTCTGCTGCTCTCTACCTCTCAAAAGAGAACTATGCTGCAAT 609
Db 779 CGAGTAGAGCGGCGATTTCTGCTGCTTCTTACTATTATAAGATAACACGCTCCATC 838
QY 610 TCCGCTGTAGCAGCCTCTCTGATTTTACACTCTGCTGCTGTAGATGATGAGCTCTTG 669
Db 839 TGCCGAGAGCAGCGCCCTGCTCGGCTCTCAGTCCCGGCTGCTAGAGCAGCAGATG 898
QY 670 TTGGAGCCATAGCCAAACAAACCCAGGAGCGAGTTTATGTATGTGTAGACACAGA 729
Db 899 CTCAGGCAATAGGAAGCAATCCAGGAAGTGAATCGTTTATGCTGTGACCCCGG 958
QY 730 CCAA 734
Db 959 CCTAA 963
```

RESULT 15

```
BC044359
LOCUS
DEFINITION
  Danio rerio, Similar to myotubularin related protein 6, clone
  MGC:55289 IMAGE:2600156, mRNA, complete cds.
ACCESSION
  BC044359
VERSION
  BC044359.1 GI:27882482
KEYWORDS
  MGC.
SOURCE
  Danio rerio (zebrafish)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.
  1 (bases 1 to 2754)
  Strausberg, R.
  Direct Submission
  Submitted (17-JAN-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
```

Email: cgpbs@mail.nih.gov
Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 101 Row: a Column: 20.

FEATURES

source

```
1..2754
  /organism="Danio rerio"
  /mol_type="mRNA"
  /db_xref="taxon:7955"
  /clone="MGC:55289 IMAGE:2600156"
  /tissue_type="whole body, adult, (one male and one female,
  including unfertilized eggs)"
  /clone_lib="Sugano_Kawakami_zebrafish_DRA"
  /lab_host="DH10B"
  /note="Vector: pME18S-FL3"
  195..2126
    /codon_start=1
    /product="similar to myotubularin related protein 6"
    /protein_id="AAH44359.1"
    /translation="MEHRTIPKVEQKLLDRFSNKSVTGTLTTLTATHLIFVNNSTSS
    QBIWTHHHIAPVEKLSLTTCPLLIQCRNFRVHVFPVPRERDCHDIYSSLLRLRP
    VSYDLYAFSNPKONDQREEGMOLIDAAEFERMGVPCDOMOLTDVNRDVKVCRTY
    EVSPFTQFLECVWLTQFPOAFSEFWFLIQHEHISCFQFNGLNSQREBLQ
    LKERTYSLWHLHSKONYLNPVSPAESOPVLEPSTQAVHFKFRNMYHQYDRSL
    HPRQSLKNVLTLSKESNELSGTQALATKQNLGVSTPLKPTPPRERSLPPRPS
    LILGLALGRKDQDEDEMFEEAREDDPPALTGGERIVESSATENAYTELPGSFG
    SRNEPAMVTLFEGVARMTC"
    748 a 628 c 694 t
    748 a 628 c 694 g
```

CDS

```
Query Match
Best Local Similarity 12.7%; Score 238.2; DB 5; Length 2754;
Matches 424; Conservative 0; Mismatches 288; Indels 3; Gaps 1;
```

```
QY 23 CGAGGTAGAAAACGTAATGGTGGATCGTTATGTAGTAAGAAACCCAGCTAATGGGA 82
Db 214 CAAAGGTAGAACAGTTAAGTCTGGATCGCT---TTAGCAACAAGTCGGTCACAGGA 270
QY 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 142
Db 271 CGCTGATCTGAGCGCCACACATTTGATCTTCTGGAGAATAATTCAACAGCTCCACGG 330
QY 143 AAACATGATTCACCTCCATCAGATTCACCTGTCGAGAGAGTTACCCATCAGTACCTGG 202
Db 331 AGATATGGATCTGTCATCACCACATCGCCCGCTGGAGAAGCTGTCCCTCAACAACCTG 390
QY 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTGGCCCACTTTGTTTAGATTCTG 262
Db 391 GCTGCCCATCTCCTCATTCAGTCCGAAATTTCCGAGTAGTCCACATTTGTGATTCGAAG 450
QY 263 ACCTTGTGTGCCATGAGGTTTATATTCTACTGCTCAAGCTTTCTCAGCCAGCATTTACCTG 322
Db 451 ACGGGGACCTGCCACACATCTACAGCTCTCTGCTTAGGCTTCTGAGACCGGTGCTCCTACG 510
QY 323 AAGATCTTTATGCTTTTCTTTATATCCCAATCTCCAAAGAGATGAGGAAAGTGGAT 382
Db 511 ACGAGCTGTACGCTTTCTCTACAAACCCCAAGCAGAACCGACGACGAGAGAGAGGCT 570
QY 383 GGAAACTGATTGACCCCAATATCAGACTTTGGGGCGTATGGGAATATCCCAACAGAACTGGA 442
```

571	GGAGGTCATAGACCTCCGCGCTGAGTTTGAGCGATGGGTGTCCTATGTGACCAAGTGC	630
443	CCATAACAGATGCCAACAACTATGAGATATATGCAGCAGCTACCTCTCTGAAATAATGG	502
631	AGCTGACAGATGTTTAACGAGGCTACAAGGTGTGCGAAACATATCCGAGACTTATAG	690
503	TTCTTAATCTGTTACCTTGGGAACGGTGGTGGAACTTCAAAGTTCAAGAGTAAGAAC	562
691	TCCCCAATTACAGCCAGTAAACCCATCATAGTCGGCAGCTCCAAATTCAGAAGCAAGGC	750
563	GNGTCCCTGTGCTCTTACCTCTACAAAGAGACAAATGCTGCCATTTGCCCTGTAGCC	622
751	GCTTTCCTGCTCTGACGTAATTTCTACGAGGAAGAAGGCAGAGTGTGTGGGTGTAGTC	810
623	AGCCTCTCTCTGGATTTTACACTCGCTGTGTAGATGATAGCTCTTGTGTGGAGGCCATT	682
811	AGCCTCTTCTGGCTTCAGCGCTCGCTGTTTGAAGACGAGCACATGCTTCAGGCCATT	870
683	GCCAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTAGACACAGACCAAGAT	737
871	GCAAGCCAAATCACAAACCGCGAAATATATGTGTATGGACGCGACCTTAAGTT	925

Search completed: December 1, 2003, 12:30:45
Job time : 6742 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 1, 2003, 10:29:49 ; Search time 615 Seconds
(without alignments)
10060.858 Million cell updates/sec

Title: US-09-941-831a-6
Perfect score: 1878
Sequence: 1 catgattacgccagcttgg.....taaaaaaa 1878

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match %	Length DB ID		
1	1878	100.0	1878 10	US-09-941-831-6	Sequence 6, Appli
2	1878	100.0	1878 12	US-10-207-175-13	Sequence 13, Appl
3	1117.8	59.5	1222 10	US-09-908-193-13	Sequence 13, Appl
4	931	49.6	1069 10	US-09-908-193-11	Sequence 11, Appl
5	317	16.9	522 10	US-09-833-381-62	Sequence 62, Appl
6	108.8	5.8	2019 10	US-09-908-805B-24	Sequence 24, Appl
7	71.4	3.8	550 11	US-09-991-936-1472	Sequence 1472, Ap
8	71.2	3.8	389 10	US-09-867-701-5266	Sequence 5266, Ap
9	71.2	3.8	419 10	US-09-908-805B-26	Sequence 26, Appl
10	69.4	3.7	401 9	US-09-922-217-28	Sequence 28, Appl
11	69.4	3.7	401 10	US-09-833-263-28	Sequence 28, Appl
12	69.4	3.7	401 13	US-10-025-380-28	Sequence 28, Appl
13	67.2	3.6	3482 11	US-09-919-039-133	Sequence 133, App
14	63.4	3.4	1675 10	US-09-833-381-1155	Sequence 1155, Ap
15	61.4	3.3	450 9	US-09-864-761-10608	Sequence 10608, A
16	60.2	3.2	2361 12	US-10-128-174-11	Sequence 11, Appl

17	60.2	3.2	2370 14	US-10-164-080-1	Sequence 1, Appli
18	60.2	3.2	2370 14	US-10-299-327-1	Sequence 1, Appli
19	60.2	3.2	3516 11	US-09-866-050A-257	Sequence 257, App
20	60.2	3.2	3516 14	US-10-152-661-257	Sequence 257, App
21	55.4	2.9	6156 12	US-10-369-978-3	Sequence 3, Appli
22	55.4	2.9	24358 12	US-10-369-978-1	Sequence 1, Appli
23	53.6	2.9	449 12	US-09-814-353-1023	Sequence 1023, Ap
24	53.6	2.9	449 12	US-09-814-353-7392	Sequence 7392, Ap
25	53.6	2.9	521 12	US-09-814-353-13777	Sequence 13777, A
26	53.6	2.9	3811 14	US-10-177-293-105	Sequence 105, App
27	53.6	2.9	5352 10	US-09-833-381-1027	Sequence 1027, Ap
28	53.6	2.9	5352 10	US-09-833-381-1028	Sequence 1028, Ap
29	53	2.8	6549 12	US-10-301-822-92	Sequence 92, Appli
30	52.4	2.8	5148 12	US-10-021-571-1	Sequence 1, Appli
31	52	2.8	704 12	US-09-814-353-19659	Sequence 19659, A
32	52	2.8	1797 9	US-09-735-368-3	Sequence 3, Appli
33	51.8	2.8	1568 12	US-10-199-937-98	Sequence 98, Appl
34	51.8	2.8	2409 10	US-09-964-899-40	Sequence 40, Appl
35	51.8	2.8	3508 12	US-10-199-937-1	Sequence 1, Appli
36	51.8	2.8	4127 12	US-10-199-937-100	Sequence 100, App
37	51.8	2.8	4275 11	US-09-972-115A-5	Sequence 5, Appli
38	51.8	2.8	4406 12	US-10-199-937-106	Sequence 106, App
39	51.8	2.8	4992 12	US-10-199-937-134	Sequence 134, App
40	51.8	2.8	5002 12	US-10-199-937-132	Sequence 132, App
41	51.8	2.8	5778 14	US-10-163-587A-14	Sequence 14, Appl
42	51.8	2.8	6018 12	US-09-849-602-11	Sequence 11, Appl
43	51.6	2.7	4297 11	US-09-972-115A-3	Sequence 3, Appli
44	51.4	2.7	3984 12	US-10-199-937-3	Sequence 3, Appli
45	51.4	2.7	4134 9	US-09-841-835-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-941-831-6
; Sequence 6, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: PTO49P1
; FILE REFERENCE: Serine/Threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941,831
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-941-831-6

Query Match				100.0%	Score 1878;	DB 10;	Length 1878;
Best Local Similarity				100.0%;	Pred. No. 0;		
Matches 1878;				Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CATGATTACGCCAACGCTTGGCAGAGGGTAGAAAAACGTGAATTTGGTGCATCGTTATG	60				
Db	1	CATGATTACGCCAACGCTTGGCAGAGGGTAGAAAAACGTGAATTTGGTGCATCGTTATG	60				
Qy	61	AGTAAAGAACACGACGTAATGGGATTTCTTATCTTACTGCAACCCACCTGATCTATGTGGAG	120				
Db	61	AGTAAAGAACACGACGTAATGGGATTTCTTATCTTACTGCAACCCACCTGATCTATGTGGAG	120				
Qy	121	GCTTCAGGTGACGCCGGAAGAAAACATGGATTGACATCCATCATTGCCACTGTGGAG	180				
Db	121	GCTTCAGGTGACGCCGGAAGAAAACATGGATTGACATCCATCATTGCCACTGTGGAG	180				
Qy	181	AGTTTACCACCTACCTAGCTGGGTTGTCCCTGACCCCTCGCTGCAGAAATTTCCGGGTG	240				

```
Db 181 |||||AAGTACCACACATAGCTGGGCTGTGCCCTGACCTCGCTGCAAGAAATTTCCGGGTG 240
Qy 241 |||||GCCACTTTGTTAGATCTGACCTGTGTGTCATGAGTTTATATTTTCACTGCTCAAG 300
Db 241 |||||GCCACTTTGTTAGATCTGACCTGTGTGTCATGAGTTTATATTTTCACTGCTCAAG 300
Qy 301 |||||CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATATCCCAATCCTCA 360
Db 301 |||||CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATATCCCAATCCTCA 360
Qy 361 |||||AAAGAGATGAGGGAAGTGGATGGAACCTGATTTGACCCCAATATCAGACTTTGGGCGTATG 420
Db 361 |||||AAAGAGATGAGGGAAGTGGATGGAACCTGATTTGACCCCAATATCAGACTTTGGGCGTATG 420
Qy 421 |||||GGAAATACCAACAGAACTGGACCATTAACAGATGCCAACAGAACTATGAGATATGCAGC 480
Db 421 |||||GGAAATACCAACAGAACTGGACCATTAACAGATGCCAACAGAACTATGAGATATGCAGC 480
Qy 481 |||||ACCTACCTCCTGAAATAGTGGTTCCTAAATCTGTTACCTTGGGAACGGTGTGGAAGT 540
Db 481 |||||ACCTACCTCCTGAAATAGTGGTTCCTAAATCTGTTACCTTGGGAACGGTGTGGAAGT 540
Qy 541 |||||TCAAAGTTCAGAAGTAAAGAACTGTCCTGTGCTCTCTACCTCTACAAAGAGAAACAAT 600
Db 541 |||||TCAAAGTTCAGAAGTAAAGAACTGTCCTGTGCTCTCTACCTCTACAAAGAGAAACAAT 600
Qy 601 |||||GCTGCCATTTGCCGTGTAGCCAGCCTCTCTCTGGATTTTACACTGCTGTGTAGATGAT 660
Db 601 |||||GCTGCCATTTGCCGTGTAGCCAGCCTCTCTCTGGATTTTACACTGCTGTGTAGATGAT 660
Qy 661 |||||GAGCTCTGTGGAGGCCATTAGCCAAACAAACCCAGGGAGCCAGTTTATGTTGTA 720
Db 661 |||||GAGCTCTGTGGAGGCCATTAGCCAAACAAACCCAGGGAGCCAGTTTATGTTGTA 720
Qy 721 |||||GACACAGAACCAAGATCTGGCATTTCTTGTGCTCTAATAGAGATAGTTTCTCCAATTA 780
Db 721 |||||GACACAGAACCAAGATCTGGCATTTCTTGTGCTCTAATAGAGATAGTTTCTCCAATTA 780
Qy 781 |||||GCCAAGATGAACCTCATGACATCACCAAGATCTTCTCCCTCTGCGAGCCGACAGGAG 840
Db 781 |||||GCCAAGATGAACCTCATGACATCACCAAGATCTTCTCCCTCTGCGAGCCGACAGGAG 840
Qy 841 |||||GAGGAGACACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACACGAC 900
Db 841 |||||GAGGAGACACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACACGAC 900
Qy 901 |||||TCCTATACCTTTGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAGG 960
Db 901 |||||TCCTATACCTTTGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAGG 960
Qy 961 |||||AGTGGCTGGGGTGTCTCTGGGACACCTTTGGCGTGTGGCTCTTTATGGCCACTTTGAGC 1020
Db 961 |||||AGTGGCTGGGGTGTCTCTGGGACACCTTTGGCGTGTGGCTCTTTATGGCCACTTTGAGC 1020
Qy 1021 |||||TGTTTGCAGTCTCTTAGCCCATGTGCTGATGTTGACAGCTTGTATGTCAAGGCACAG 1080
Db 1021 |||||TGTTTGCAGTCTCTTAGCCCATGTGCTGATGTTGACAGCTTGTATGTCAAGGCACAG 1080
Qy 1081 |||||ACGCCACTTTTCACTGCTGACGTATGGCCATCTGGACCTGTGATGCTGTGCTTTTGGAA 1140
Db 1081 |||||ACGCCACTTTTCACTGCTGACGTATGGCCATCTGGACCTGTGATGCTGTGCTTTTGGAA 1140
Qy 1141 |||||GCTGGTGCCTCTCTCGTGTAGCATCTACAACTGTTTCTCCCGTGTCTACAGCTGCC 1200
Db 1141 |||||GCTGGTGCCTCTCTCGTGTAGCATCTACAACTGTTTCTCCCGTGTCTACAGCTGCC 1200
Qy 1201 |||||CGTGATGGTGTCTGCTATCTCTGAGGAGCTCTAGACCATGCTGTCAGAGGCCACGTC 1260
Db 1201 |||||CGTGATGGTGTCTGCTATCTCTGAGGAGCTCTAGACCATGCTGTCAGAGGCCACGTC 1260
Qy 1261 |||||AAAGCTAAACTACAGTCTGGGATCAAAATAGCTTCTGTCGCTCTGCTTATTTG 1320
```

```
Db 1261 |||||AAAGCTAAACTACAGTCTGGGATCAAAATAGCTTCTATGTTCTGGCCCCCTCTATTG 1320
Qy 1321 |||||GCCGAGTCTACGGGACCTGGACTGTTTCCGCTGCTTTTCTCCACGGGCGACACCT 1380
Db 1321 |||||GCCGAGTCTACGGGACCTGGACTGTTTCCGCTGCTTTTCTCCACGGGCGACACCT 1380
Qy 1381 |||||GACTACAACTGACCTGACCCAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTT 1440
Db 1381 |||||GACTACAACTGACCTGACCCAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTT 1440
Qy 1441 |||||GAAATCTGCTCCATCAATAATTGTGAGCAGAGTATATCCAGCTGTTAATCGATTTGGT 1500
Db 1441 |||||GAAATCTGCTCCATCAATAATTGTGAGCAGAGTATATCCAGCTGTTAATCGATTTGGT 1500
Qy 1501 |||||GCTAATATCTACTTCCATCTCTCTCCCTTGGACCTCAAGATGATAAAGCAAT 1560
Db 1501 |||||GCTAATATCTACTTCCATCTCTCTCCCTTGGACCTCAAGATGATAAAGCAAT 1560
Qy 1561 |||||GCATTGCTCTACAGGCGGAGCCACTCCAGGTCACCTTCTATCAAGGTCGGTTTGTGTC 1620
Db 1561 |||||GCATTGCTCTACAGGCGGAGCCACTCCAGGTCACCTTCTATCAAGGTCGGTTTGTGTC 1620
Qy 1621 |||||GTCCGAGAGCCTTGTGCGAGCTGCGCCAGCCACCAAGCCATCAACAGCTGGATATTCCT 1680
Db 1621 |||||GTCCGAGAGCCTTGTGCGAGCTGCGCCAGCCACCAAGCCATCAACAGCTGGATATTCCT 1680
Qy 1681 |||||CCCATGTTGATTAGCTACCTAAACACCACTGTTAATCTTGCACTCTCCAGGACTTCA 1740
Db 1681 |||||CCCATGTTGATTAGCTACCTAAACACCACTGTTAATCTTGCACTCTCCAGGACTTCA 1740
Qy 1741 |||||TGATGCTCCGAAACACCTGCGGACTCACCTAGCTGGAGAGCATTCACAGCTCATCCA 1800
Db 1741 |||||TGATGCTCCGAAACACCTGCGGACTCACCTAGCTGGAGAGCATTCACAGCTCATCCA 1800
Qy 1801 |||||CTTACCTGGAGCTGCTCTCTGATTTATCTCCCAATAAAATTTCTCCAGAAAATAAGTA 1860
Db 1801 |||||CTTACCTGGAGCTGCTCTCTGATTTATCTCCCAATAAAATTTCTCCAGAAAATAAGTA 1860
Qy 1861 |||||AAAAAAAAAAAAAAAAAAAA 1878
Db 1861 |||||AAAAAAAAAAAAAAAAAAAA 1878
```

RESULT 2

```
US-10-207-175-13
; Sequence 13, Application US/10207175
; Publication No. US20030157508A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ/4P1
; CURRENT APPLICATION NUMBER: US/10/207,175
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: PCT/US01/01395
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-207-175-13
```

Query Match 100.0%; Score 1878; DB 12; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGATTACCCAGCTTGGCAAGAGGTGAGAAACCTGAAATTTGGTGGATCCTTATGTG 60
DB 1 CATGATTACCCAGCTTGGCAAGAGGTGAGAAACCTGAAATTTGGTGGATCCTTATGTG 60
QY 61 AGTAAGAACACGCTAATGGGATCTTTATCTTACTGCAACCACTGATCTATGTGGAG 120
DB 61 AGTAAGAACACGCTAATGGGATCTTTATCTTACTGCAACCACTGATCTATGTGGAG 120
QY 121 GCTTCAGGTGACGCCGGAAGAAACATGGATTGCACTCCATCACAATTCGCACTGTGGAG 180
DB 121 GCTTCAGGTGACGCCGGAAGAAACATGGATTGCACTCCATCACAATTCGCACTGTGGAG 180
QY 181 AGTTACCCATCACTAGCTGGGTGTCCCTGACCCCTCGCTGCAAGAAATTTCCGGGTG 240
DB 181 AGTTACCCATCACTAGCTGGGTGTCCCTGACCCCTCGCTGCAAGAAATTTCCGGGTG 240
QY 241 GCCCACTTTTGTAGATTCTGACCTTGTGTCGCAATGAGTTTATATTTCACTGCTCAAG 300
DB 241 GCCCACTTTTGTAGATTCTGACCTTGTGTCGCAATGAGTTTATATTTCACTGCTCAAG 300
QY 301 CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATCCCAATCCTCA 360
DB 301 CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATCCCAATCCTCA 360
QY 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTATG 420
DB 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTATG 420
QY 421 GGAATACCCACAGAACTGGACCATTAACAGATGCCAACAGAACTATGAGATATCAGC 480
DB 421 GGAATACCCACAGAACTGGACCATTAACAGATGCCAACAGAACTATGAGATATCAGC 480
QY 481 ACCTACCTCTCAATAGTGGTTCTTAATCTGTTACCTTGGGAACGGTGTGGAAT 540
DB 481 ACCTACCTCTCAATAGTGGTTCTTAATCTGTTACCTTGGGAACGGTGTGGAAT 540
QY 541 TCAAGTTCAGAAAGTGAAGACCTGCTCCCTGCTCTCTTACCTCTACAAAGAGAAAT 600
DB 541 TCAAGTTCAGAAAGTGAAGACCTGCTCCCTGCTCTCTTACCTCTACAAAGAGAAAT 600
QY 601 GCTGCCATTTGCGCTGTAGCCAGCTCTCTCTGGAATTTTACACTGCTGTGTAGAT 660
DB 601 GCTGCCATTTGCGCTGTAGCCAGCTCTCTCTGGAATTTTACACTGCTGTGTAGAT 660
QY 661 GAGCTCTGTGTGAGGCCATTAGCCAAACAAACCCAGGGAGCCAGTTATGATGTTA 720
DB 661 GAGCTCTGTGTGAGGCCATTAGCCAAACAAACCCAGGGAGCCAGTTATGATGTTA 720
QY 721 GACAAGACCAAGATCTGGCAATTTCTTGTGCTCATATGAGAAATAGTTCTCCAATTA 780
DB 721 GACAAGACCAAGATCTGGCAATTTCTTGTGCTCATATGAGAAATAGTTCTCCAATTA 780
QY 781 GCCAAGATGAACCTCATGACATCAAGATCTTCTCCTCTGAGCCCGCAAGAGGAG 840
DB 781 GCCAAGATGAACCTCATGACATCAAGATCTTCTCCTCTGAGCCCGCAAGAGGAG 840
QY 841 GAGGAGACACTGACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAGGAC 900
DB 841 GAGGAGACACTGACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAGGAC 900
QY 901 TCCTATCTTTGGACCACTTTTGGCCGAGGAGCGTTACAAACGTTTTCATCAACAGCAG 960
DB 901 TCCTATCTTTGGACCACTTTTGGCCGAGGAGCGTTACAAACGTTTTCATCAACAGCAG 960
QY 961 AGTGGCTGGGTGTTCTGGGACACCCCTTGGCTTCTTATGGCCACTTGGAGC 1020
DB 961 AGTGGCTGGGTGTTCTGGGACACCCCTTGGCTTCTTATGGCCACTTGGAGC 1020
QY 1021 TGTTCGAACTCTCTAGCCCATGCTGATGTTGACAGCTTGTGATGTCAGGACAG 1080
DB 1021 TGTTCGAACTCTCTAGCCCATGCTGATGTTGACAGCTTGTGATGTCAGGACAG 1080
QY 1081 ACGCCACTTTTCACTGCTGTCAGTCATGGCCATCTGGAATCTGACGTGTGCTTTTGGAA 1140

DB 1081 ACGCCACTTTTCACTGCTGTCAGTCATGGCCATCTGGAATCTGCTGCTTTTGGAA 1140
QY 1141 GCTGGTGCCTCTCTGGTGGTAGCATCTACAACTGTTCTCCGCTGCTACAGTGGC 1200
DB 1141 GCTGGTGCCTCTCTGGTGGTAGCATCTACAACTGTTCTCCGCTGCTACAGTGGC 1200
QY 1201 CGTGTAGTGTGCTGTTGCTATCTCGAGGAGCTCTAGACCATGTTGTCAGAGGCCAACGTC 1260
DB 1201 CGTGTAGTGTGCTGTTGCTATCTCGAGGAGCTCTAGACCATGTTGTCAGAGGCCAACGTC 1260
QY 1261 AAAGCTAAACTACAGTCTGGGCATCAAACTAGTCTTCAATGTTCTGGCCCCCTCTATTG 1320
DB 1261 AAAGCTAAACTACAGTCTGGGCATCAAACTAGTCTTCAATGTTCTGGCCCCCTCTATTG 1320
QY 1321 GCCGAGTCTACGGGACCTGCACTGTTTCCGCTGCTTTTCTCCAGGGGAGACCT 1380
DB 1321 GCCGAGTCTACGGGACCTGCACTGTTTCCGCTGCTTTTCTCCAGGGGAGACCT 1380
QY 1381 GACTACACTGCACTGACCCAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTT 1440
DB 1381 GACTACACTGCACTGACCCAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTT 1440
QY 1441 GAAATCTGCTCCATCAATAATTTGAGCCAGAGTATATCCAGCTGTTAATCGATTTG 1500
DB 1441 GAAATCTGCTCCATCAATAATTTGAGCCAGAGTATATCCAGCTGTTAATCGATTTG 1500
QY 1501 GCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAT 1560
DB 1501 GCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAT 1560
QY 1561 GCAATGCTGTACAGCCGAGCCACTCCACGGTCACTTCTATCAAGCTCGTTTAGTC 1620
DB 1561 GCAATGCTGTACAGCCGAGCCACTCCACGGTCACTTCTATCAAGCTCGTTTAGTC 1620
QY 1621 GTCCGAGAGCCTTGTGCCAGCTGCGCCAGCCACCAAGCCATCAACAGCTGGATTTCT 1680
DB 1621 GTCCGAGAGCCTTGTGCCAGCTGCGCCAGCCACCAAGCCATCAACAGCTGGATTTCT 1680
QY 1681 CCCATGTTGATGATCTACCTAAACACCACTGTTAATCTTGGAGTCTCCAGGAACTTA 1740
DB 1681 CCCATGTTGATGATCTACCTAAACACCACTGTTAATCTTGGAGTCTCCAGGAACTTA 1740
QY 1741 TCATGCTCCGAAACCACTGCGGAGCTCACCTGAGAGCATTAAGCTCATCCA 1800
DB 1741 TCATGCTCCGAAACCACTGCGGAGCTCACCTGAGAGCATTAAGCTCATCCA 1800
QY 1801 CTTACTGAGCTGCTCTCTCTATTTATCTCCCAATAAATTTCTCCAGAAATAAGTA 1860
DB 1801 CTTACTGAGCTGCTCTCTCTATTTATCTCCCAATAAATTTCTCCAGAAATAAGTA 1860
QY 1861 AAAAAAAAAAAAAAAAAA 1878
DB 1861 AAAAAAAAAAAAAAAAAA 1878

RESULT 3

US-09-908-193-13
; Sequence 13, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHINKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADICARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650

; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicolas A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 10976Z
; CURRENT APPLICATION NUMBER: US/09/908,805B
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (2000)
; OTHER INFORMATION: n is unsure
US-09-908-805B-24.

Query Match 5.8%; Score 108.8; DB 10; Length 2019;
Best Local Similarity 54.1%; Pred. No. 1.2e-23;
Matches 302; Conservative 1; Mismatches 228; Indels 27; Gaps 3;
QY 874 CTCATCAACGACGTGTATGACACGACGAGTGGCTGG-----GGTGTCTCTGGGACA 984
Db 167 CTCATGATGACGACCTATGAGGGGACCTCCAGACCCCTCAGAACCTACTGCAAGAGGAG 226.
QY 934 CGTTACAAACGTTTTCATCAACACGACGAGTGGCTGG-----GGTGTCTCTGGGACA 984
Db 227 AGTACCGGAGCGCATCATGAGAGTCTGTCTGTGTGCGGTGGCTTCCCTGCACA 286
QY 985 CCCTTGGCTGTGCTGCTTCTTATGACCACTTGAGCTGTGTTGCAAGTCTCTTTCAGCCCAT 1044
Db 287 CCACTGAGGATCGCAGCCACTGCGAGCCATGGGAACCTGTGTGACTTCCCTCATACGAAA 346
QY 1045 GGTGCTGATGTTACAGCTTGGATGTCAGGACAGACGCCACTTTTCTACTGCTGTCAGT 1104
Db 347 GGGGCGGAGTGGACCTGTGTGGATGTCAGGGGCGACACTGCCCTGTATGTGGCTGTAGTG 406
QY 1105 CATGGCCATCTGACTGTGTACGTGTCTTTTGGAGCTGGTGCCTCTCTCTGTGTGTAGC 1164
Db 407 AACGGGCACTTGGAGACACTGAGATCCCTTTTGGAGCTGGTGTCTGATCCCAAGCGCAGC 466
QY 1165 ATCTACAAACACTGTTTCTCCCGTGTCTCACAGCTGCCCGTGTGATGCTGTGTATCTCTG 1224
Db 467 CGGCAACACCGCAGCACTCTGTGTACCATGCTCTGTGTGGTAGGACGACATCTCTG 526
QY 1225 CAGGAGCTCTAGACCATGCTGTGACAGGCGCAAGCTCAAGCTAACT-----ACC 1274
Db 527 AAGGCTCTTATCAGGTATGGGGCAGATGTTGATGTCAACCATCATCTGAAATCTGCACAC 586
QY 1275 AGTCTGGGCATCAACATAGCTTCATGTTCT-----GGCCCCCTCTATTGTCGCGCA 1326
Db 587 CGGCCCCCTTTTTCAGCGGGCTAACTCTCTGGTGTGTCTCTCTATACATCACTGCT 646
QY 1327 GTCTACGGGCACTGGAAGCTGTTTCCCGCTGTTTGTCTCCAGGGGCGACCCCTGACTAC 1386
Db 647 GCCTACCAACAACTTCTCAGTGTCTTTCAGGCTGCTCTTGCAGGCTGGGGCAAACTCTGACTTC 706
QY 1387 AACTGCACTGACGAGGC 1404
Db 707 AATTGCAATGGCCCTGTC 724

RESULT 7
US-09-991-936-1472
; Sequence 1472, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.

; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1472
; LENGTH: 550
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1472

Query Match 3.8%; Score 71.4; DB 11; Length 550;
Best Local Similarity 49.9%; Pred. No. 6.9e-12;
Matches 180; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 129 TGCAGCCCGGAAAGAAACATGATTGCACTCCCATCACATTCGACATTCGAGAGGTTACC 188
Db 29 TATCGGAAATGAAACATGATATACACACATATTCGAGCAGTAGAAGATTACC 88
QY 189 CATCACTAGCCTGGGTGTCCTGACCTCCGCTGCAAGAAATTTCCGGGTGGCCCACTT 248
Db 89 GTTGAGTACTACAGGATGCCACTTCAAAATTCGATGCAAAACATTCCAATCAGTAATAT 148
QY 249 TCTTTTAGATTCGACCTGTGTGCGCATGAGCTTATATTCACCTGCTCAAGCTTCTCA 308
Db 149 TCTTAAAGAAAGAAACGATAGCCATGACGTTTATATATCTTTGTTCGAATTAGCCCA 208
QY 309 GCAGCATTACCTGAAGATCTTTATGCTTTTCTTATATCCCAATCTCAAAAGAGAT 368
Db 209 ACTGTTTCTATCGATGATCTGTTTCTCAGTACCATATAAATAAGATATGCA 268
QY 369 GAGGAAAGTGGATGGAACCTGATGACCAATATACAGCTTTGGGCGTATGGGAATACC 428
Db 269 TAAGAGCGAAGTTTGAATTTATTTAGAGGATGAATTTAAAGAAATGAATGTTCC 328
QY 429 CAAAGAACTGACCATACAGATGCCCAAGAACTATGAGATATGAGACACCTACCC 488
Db 329 AAATGATGATGGATTTTACACAGATTTAAATGAAATTTATGAGCTATGTATCTATCC 388
QY 489 T 489
Db 389 T 389

RESULT 8
US-09-867-701-5266
; Sequence 5266, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5266
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-5266

GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(401)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-263-28

Query Match 3.7%; Score 69.4; DB 10; Length 401;
Best Local Similarity 52.4%; Pred. No. 2.5e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGAAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGGCGTAT 419
Db 52 AGAAAAGTTTAACTGGATGGAGACAGTTTCAATCCAGTGGAGAAATACAGGAGGCA 111
QY 420 GGGAAATACCCACAGAAACTGGACCATACAGATGCCAAGAACTATGAGATATGCAG 479
Db 112 GGGCTTGGCCCAATCACCATTGGAGATAACTTTTATTATAAGTCTATGAGCTCTGCGA 171
QY 480 CACCTACCTCTGAAATAGTGTCTCTAAATCTGTACCTTGGGAACGGTGTGGAAG 539
Db 172 CACTTACCTCTCTTTTGGTGGTTCGGTATCGTGCTCANATGATGACCTCCGGAGAGT 231
QY 540 TTCAAAGTTTCAGAAAGTAAAGAACTGTCCTGTGCTCTCTACCTCTACAAAGAGACAA 599
Db 232 TGCACCTTTAGTTCGGAAATCGAATTCAGTGTGCTATGAGTTCATCCAGAAATAA 291
QY 600 TGTGCACTTGGCGTGTAGCCAGCCTCTCTCTGGATT---TTACACTGCGTGTGTAGA 656
Db 292 GACGGTCATTGTCGGTTCAGTCCAGCTCTTGTTCGGTATGATGGGAAACGAATAAGA 351
QY 657 TGATGAGCTCTTTGGAGGCCATTAGCCAAACAAA 692
Db 352 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 387

RESULT 12
US-10-025-380-28
; Sequence 28, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 212
; OTHER INFORMATION: n = A,T,C or G
US-10-025-380-28

Query Match 3.7%; Score 69.4; DB 13; Length 401;
Best Local Similarity 52.4%; Pred. No. 2.5e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGAAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGGCGTAT 419
Db 52 AGAAAAGTTTAACTGGATGGAGACAGTTTCAATCCAGTGGAGAAATACAGGAGGCA 111
QY 420 GGGAAATACCCACAGAAACTGGACCATACAGATGCCAAGAACTATGAGATATGCAG 479
Db 112 GGGCTTGGCCCAATCACCATTGGAGATAACTTTTATTATAAGTCTATGAGCTCTGCGA 171
QY 480 CACCTACCTCTGAAATAGTGTCTCTAAATCTGTACCTTGGGAACGGTGTGGAAG 539
Db 172 CACTTACCTCTCTTTTGGTGGTTCGGTATCGTGCTCANATGATGACCTCCGGAGAGT 231
QY 540 TTCAAAGTTTCAGAAAGTAAAGAACTGTCCTGTGCTCTCTACCTCTACAAAGAGACAA 599
Db 232 TGCACCTTTAGTTCGGAAATCGAATTCAGTGTGCTATGAGTTCATCCAGAAATAA 291
QY 600 TGTGCACTTGGCGTGTAGCCAGCCTCTCTCTGGATT---TTACACTGCGTGTGTAGA 656
Db 292 GACGGTCATTGTCGGTTCAGTCCAGCTCTTGTTCGGTATGATGGGAAACGAATAAGA 351
QY 657 TGATGAGCTCTTTGGAGGCCATTAGCCAAACAAA 692
Db 352 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 387

RESULT 13
US-09-919-039-133
; Sequence 133, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 133
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 1823159CB1
US-09-919-039-133

Query Match 3.6%; Score 67.2; DB 11; Length 3482;
Best Local Similarity 52.1%; Pred. No. 5.7e-10;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGAAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGGCGTAT 419

Search completed: December 1, 2003, 13:44:05
Job time : 620 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:53 ; Search time 68 Seconds
(without alignments)
1048.062 Million cell updates/sec

Title: US-09-941-831A-20

Perfect score: 2343

Sequence: 1 MRESGWKLIDPISDFGRMGI.....QAINQLDIPPLISYLKQL 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq 15Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2343	100.0	449	22	Human serine/threose
2	2343	100.0	449	22	Human HIBJ89 seri
3	2343	100.0	449	22	Human secreted pro
4	2331	99.5	580	22	Human secreted pro
5	1890	72.1	360	23	Human NOV5b protei
6	1635	69.8	318	23	Human polypeptide
7	1629	69.5	351	23	Human NOV5a protei
8	642.5	27.4	370	23	Novel human protei
9	499	21.3	337	24	Human cytoskeleton

10	499	21.3	341	22	AAU20362	Human secreted pro
11	499	21.3	343	22	AAU20580	Human secreted pro
12	480	20.5	350	19	AAW62621	Mus musculus SOCS7
13	423	18.1	629	22	AAW73229	Human phosphatase
14	392.5	16.8	482	23	ABP62857	Human polypeptide
15	360	15.4	639	22	ABB65465	Drosophila melanog
16	360	15.4	676	22	ABB59964	Drosophila melanog
17	360	15.4	676	22	ABB66818	Drosophila melanog
18	297	12.7	393	22	AAW93364	Human protein sequ
19	281	12.0	665	22	AAW78876	Human protein sequ
20	281	12.0	673	23	AAE35082	Human kinase and p
21	275.5	11.8	858	22	ABG09695	Novel human diagno
22	273	11.7	705	22	AAW79860	Human protein sequ
23	270.5	11.5	278	21	AAW53886	A suppressor of cy
24	264.5	11.3	278	22	AAW95322	Human protein sequ
25	264	11.3	621	18	AAW36451	Human tyrosine pho
26	261.5	11.2	276	22	AAW39002	Human polypeptide
27	261.5	11.2	329	23	AAW50585	Benign prostatic h
28	256.5	10.9	299	23	ABP41229	Human ovarian anti
29	246	10.5	323	23	ABB05439	Human C/EBP-1 pr
30	244.5	10.4	690	22	ABB11804	Human ASB-3 protei
31	243.5	10.4	518	21	AAW53889	A suppressor of cy
32	243.5	10.4	518	22	AAW92538	Human protein sequ
33	241	10.3	342	22	ABB11919	Human cytokine sig
34	241	10.3	342	22	AAW79442	Human protein sequ
35	235	10.0	407	23	ABP69199	Human polypeptide
36	235	10.0	435	22	AAW66710	Human cell growth
37	235	10.0	435	22	AAW93879	Human protein sequ
38	235	10.0	435	23	ABP69654	Human polypeptide
39	235	10.0	627	23	AAO17136	Human cancer cell
40	233.5	10.0	1377	22	ABG08072	Novel human diagno
41	231	9.9	362	21	AAW51508	Human secreted pro
42	230	9.8	293	21	AAW51425	Human secreted pro
43	230	9.8	293	21	AAW51427	Human secreted pro
44	228	9.7	784	22	ABB62206	Drosophila melanog
45	224.5	9.6	447	22	ABU53030	Cell structure and

ALIGNMENTS

RESULT 1

AAW85477

ID AAW85477 standard; Protein; 449 AA.

XX AAW85477;

AC AAW85477;

XX 25-SEP-2001 (first entry)

DT Human serine/threonine phosphatase (clone ID HIBJ89).

DE Serine/threonine phosphatase; neotropic; neuroprotective; cytostatic;
KW dermatological; immunosuppressive; antiinflammatory; antibacterial;
KW anti-HIV; antiparkinsonian; antitickling; antianemic; antiarthritic;
KW antirheumatic; virucide; hepatotropic; cerebrotropic; vulnerary;
KW antiinflammatory; nephrotropic; gene therapy; vaccine.

XX Homo sapiens.

OS WO200155388-A1.

PN 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01395.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 02-MAR-2000; 2000US-0186350.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Barash SC, Ruben SM;

PI

XX

DR WPI; 2001-476208/51.
DR N-PSDB; AAH46857.
XX Novel proteins of serine/threonine phosphatase family, useful for
PT diagnosing, treating, preventing and/or prognosing disorders related to
PT the proteins, including cancer, immune response and neuronal disorders
PT
XX
XX Claim 11; Page 432-433; 441pp; English.
PS
XX The invention provides serine/threonine phosphatase family polypeptides
CC and polynucleotides encoding them. The polypeptides can be expressed by
CC standard recombinant methodology. The polypeptides, polynucleotides and
CC antibodies are useful for diagnosis, prognosis, prevention and treatment
CC of neurodegenerative disorders, immune system disorders, autoimmune
CC diseases, allergic reactions, infectious diseases, hyperproliferative
CC disorders, renal disorders, cardiovascular disorders, cerebrovascular
CC disorders, respiratory disorders, endocrine disorders, gastrointestinal
CC disorders and also muscular, reproductive disorders (see AAH46846 for a
CC detailed description of the diseases that can be treated). They are also
CC useful as immune system enhancers, immunosuppressive agents, stimulator
CC of B-cell responsiveness to pathogens, activator of T-cells, to induce
CC higher affinity antibodies, as an agent to increase serum immunoglobulin
CC concentrations, to accelerate recovery of immunocompromised individuals,
CC to boost immunoresponsiveness in aged populations and/or neonates, as
CC regulator of antigen presentation, as a means to induce tumour
CC proliferation in pathologies e.g. AIDS, and/or common variable immuno
CC deficiency. The polypeptides and polynucleotides are useful to prevent
CC skin aging, for preventing hair loss, to stimulate growth and
CC differentiation of hemopoietic cells and bone marrow cells, for
CC supporting cell culture of primary tissues, to modulate mammalian
CC characteristics such as body weight, height, eye color, hair color and
CC skin, to modulate mammalian metabolism to change a mammal's mental or
CC physical state, and as food additive or preservative. The present
CC sequence represents a human serine/threonine phosphatase.
XX
SQ Sequence 449 AA;
Query Match 100.0%; Score 2343; DB 22; Length 449;
Best Local Similarity 100.0%; Pred. No. 2.6e-252;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRESGKGLDIPISDFGRMGIPNRRNWTITDANRNYEICSTYPPPIVVPKSVTLGTGVGSSK 60
DB 1 MRESGKGLDIPISDFGRMGIPNRRNWTITDANRNYEICSTYPPPIVVPKSVTLGTGVGSSK 60
QY 61 FRSKRVPVLSYLYKENNAICRCSQPLSGFYTRCVDDLELLLEAISQTNPGSQFMVYVDT 120
DB 61 FRSKRVPVLSYLYKENNAICRCSQPLSGFYTRCVDDLELLLEAISQTNPGSQFMVYVDT 120
QY 121 RPKIWHFLVLMRIVLQAKNLMIDTKIFSLLOPKDEEDTTEKQALNQAVYNDYSY 180
DB 121 RPKIWHFLVLMRIVLQAKNLMIDTKIFSLLOPKDEEDTTEKQALNQAVYNDYSY 180
QY 181 TLIDQLLRQERYKRFINSRSGWVGPTPLRLAASGYHLSCLQVLHGADVDSLDVKAQTP 240
DB 181 TLIDQLLRQERYKRFINSRSGWVGPTPLRLAASGYHLSCLQVLHGADVDSLDVKAQTP 240
QY 241 LFTAVSHGLDCVRVLEAGAPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEANVKA 300
DB 241 LFTAVSHGLDCVRVLEAGAPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEANVKA 300
QY 301 KLPVWASNTASCSGPYLAAYGHLDPCFRLLLLHGADPDYNCCTDQGLLARVPRPTLLBI 360
DB 301 KLPVWASNTASCSGPYLAAYGHLDPCFRLLLLHGADPDYNCCTDQGLLARVPRPTLLBI 360
QY 361 CLHHCNEPEYIQLIDFGANILPSSLTDLTSQDDKGIALLLQARATPRSLLSQVRLVVR 420
DB 361 CLHHCNEPEYIQLIDFGANILPSSLTDLTSQDDKGIALLLQARATPRSLLSQVRLVVR 420
QY 421 RALCOAQGPQAINQLDIPPMILSYLKHQL 449
DB 421 RALCOAQGPQAINQLDIPPMILSYLKHQL 449

RESULT 2
AAG77803
ID AAG77803 standard; Protein; 449 AA.
XX
AC AAG77803;
XX
DT 03-DEC-2001 (first entry)
XX
DE Human HIBCU89 serine/threonine phosphatase protein sequence.
XX
KW Human; HIBCU89; serine/threonine phosphatase; PSPase; vaccine;
KW gene therapy; PSPase expression; PSPase expression; PSPase modulation;
KW immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome;
KW Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis;
KW inflammation; Crohn's disease; inflammatory bowel disease; appendicitis;
KW rheumatoid arthritis; cellular proliferative disorder; lymphoma;
KW lung cancer; intestinal cancer; cardiovascular disorder; aneurysm;
KW Scimitar syndrome; Ebstein's anomaly.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 29..35 /note= "Immunogenic region 1"
FT Region 59..66 /note= "Immunogenic region 2"
FT Region 106..112 /note= "Immunogenic region 3"
FT Region 154..169 /note= "Immunogenic region 4"
FT Region 175..180 /note= "Immunogenic region 5"
FT Region 185..201 /note= "Immunogenic region 6"
FT Region 336..344 /note= "Immunogenic region 7"
FT Region 391..396 /note= "Immunogenic region 8"
XX
PN WO200164703-A1.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06256.
XX
PR 02-MAR-2000; 2000US-0186350.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ebner R, Ruben SM;
XX
DR WPI; 2001-530113/58.
DR N-PSDB; AAH78730.
XX
PT Nucleic acids encoding serine/threonine phosphatase polypeptides,
PT useful for preventing, diagnosing and/or treating, e.g. Crohn's
PT disease, lung cancer and Scimitar syndrome -
XX
PS Claim 11; Page 321-323; 335pp; English.
XX
CC The present sequence represents the human HIBCU89 protein which is
CC claimed in the invention. The invention comprises novel human
CC serine/threonine phosphatase (PSPase) polypeptides and polynucleotides.
CC The PSPase polynucleotides and polypeptides of the invention may be used
CC in the prevention (vaccine), diagnosis and treatment (gene therapy) of
CC diseases associated with inappropriate PSPase expression. The PSPase
CC polynucleotides of the invention may be used as DNA probes to detect and
CC quantitate the presence of similar nucleic acids in samples. The PSPase
CC polypeptides may be used as antigens in the production of antibodies
CC against the PSPase polypeptides and in assays to identify modulators of
CC PSPase expression and activity. The anti-PSPase antibodies and

antagonists may also be used to down regulate expression and activity, the anti-PPase antibodies may also be used as diagnostic agents for detecting the presence of pspase polypeptides in samples. Disorders that may be prevented, diagnosed and/or treated by the invention are: immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome, Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory bowel disease, appendicitis and rheumatoid arthritis); cellular proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and aneurysm).

XX SQ Sequence 449 AA;

Query Match 100.0%; Score 2343; DB 22; Length 449;

Best Local Similarity 100.0%; Pred. No. 2.6e-252;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRESGKLIIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
DB 1 MRESGKLIIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
QY 61 FRSKERPVLVLYKNNAAICRCQPSLGSFYTRCVDDLELLEAISQTNPGSQFMVVDVT 120
DB 61 FRSKERPVLVLYKNNAAICRCQPSLGSFYTRCVDDLELLEAISQTNPGSQFMVVDVT 120
QY 121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNDY 180
DB 121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNDY 180
QY 181 TLQDLRLQERYKRFINSRSGMGVPGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTP 240
DB 181 TLQDLRLQERYKRFINSRSGMGVPGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTP 240
QY 241 LFTAVSHGLDCVRVLEAGAPGGSIYNNCSFVLPAARDGAVAILQELLDHGAENYKA 300
DB 241 LFTAVSHGLDCVRVLEAGAPGGSIYNNCSFVLPAARDGAVAILQELLDHGAENYKA 300
QY 301 KLPVWASNTASCGPLVLAAYVGHLDICFRLLLLHGHADPDYNTDQGLLARVPRPTLLBI 360
DB 301 KLPVWASNTASCGPLVLAAYVGHLDICFRLLLLHGHADPDYNTDQGLLARVPRPTLLBI 360
QY 361 CLHNCPEYIQLLIDFGANILPSSLDLTSDQDKGIALLLQARATPRSLLSQVRLVVR 420
DB 361 CLHNCPEYIQLLIDFGANILPSSLDLTSDQDKGIALLLQARATPRSLLSQVRLVVR 420
QY 421 RALCQAGQPOAINQLDIPMLISYLKHQL 449
DB 421 RALCQAGQPOAINQLDIPMLISYLKHQL 449

RESULT 3

AAU20663

ID AAU20663 standard; Protein; 449 AA.

XX AC AAU20663;

XX AC AAU20663;

DT 04-DEC-2001 (first entry)

XX XX Human secreted protein, Seq ID No 655.

XX XX Human secreted protein, Seq ID No 655.

XX KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.

XX OS Homo sapiens.

XX XX

PN WO200155326-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01347.

XX PR 31-JAN-2000; 2000US-0179065.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Barash SC, Ruben SM;

XX DR WPI; 2001-451931/48.

XX DR N-PSDB; AAS33372.

XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing

XX PT or treating medical conditions -

XX PS Claim 11; SEQ ID No 655; 753pp; English.

XX CC The invention relates to novel isolated nucleic acid molecules (I)

XX CC encoding human secreted proteins (II). (I) and (II) are used to prevent,

XX CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in

XX CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in

XX CC the prevention, treatment and diagnosis of diseases associated with

XX CC inappropriate expression of secreted proteins. (I) and complementary

XX CC sequences may also be used as DNA probes in diagnostic assays (e.g.

XX CC polymerase chain reactions (PCR)) to detect and quantitate the presence

XX CC of similar nucleic acid sequences in samples, and so which patients may

XX CC be in need of restorative therapy. (II) may also be used as antigens in

XX CC the production of antibodies and in assays to identify modulators

XX CC (agonists and antagonists) of the expression and activity of the secreted

XX CC proteins. The anti-(II) antibodies and antagonists may also be used to

XX CC down regulate expression and activity of (II). The anti-(II) antibodies

XX CC may also be used as diagnostic agents for detecting the presence of (II)

XX CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The

XX CC disorders include for example: immune/autoimmune diseases (e.g. HIV

XX CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis

XX CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.

XX CC melanomas, neoplasms of the breast or liver, Sezary syndrome and

XX CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,

XX CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/

XX CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,

XX CC angina and thrombosis), infections caused by bacteria, viruses and

XX CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),

XX CC agonists, antagonists and antibodies can also be used to promote wound

XX CC healing, maintain organs before transplantation, and support cell culture

XX CC of primary tissues. AAU20342-AAU20666 represent human secreted protein

XX CC amino acid sequences, and related sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed

XX CC specification but was obtained in electronic format directly from WIPO

XX CC at: ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 449 AA;

Query Match 100.0%; Score 2343; DB 22; Length 449;

Best Local Similarity 100.0%; Pred. No. 2.6e-252;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRESGKLIIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60

DB 1 MRESGKLIIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60

QY 61 FRSKERPVLVLYKNNAAICRCQPSLGSFYTRCVDDLELLEAISQTNPGSQFMVVDVT 120

DB 61 FRSKERPVLVLYKNNAAICRCQPSLGSFYTRCVDDLELLEAISQTNPGSQFMVVDVT 120

QY 121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNDY 180

DB 121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNDY 180

QY 181 TLQDLRLQERYKRFINSRSGMGVPGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTP 240

DB 181 TLQDLRLQERYKRFINSRSGMGVPGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTP 240

QY 241 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 300
 DB 241 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 300
 QY 301 KLPWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 360
 DB 301 KLPWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 360
 QY 361 CLHNCPEPIYQLIDFGANIYLPISLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 420
 DB 361 CLHNCPEPIYQLIDFGANIYLPISLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 420
 QY 421 RALCOAGQQAQINQOLDIPPMILSYLKHQL 449
 DB 421 RALCOAGQQAQINQOLDIPPMILSYLKHQL 449
 RESULT 4
 AAU20555
 ID AAU20555 standard; Protein; 580 AA.
 AC AAU20555;
 XX 06-DEC-2001 (first entry)
 XX Human secreted protein, Seq ID No 547.
 XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
 KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
 KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
 KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
 KW Gaucher's disease; neurological disease; cerebrovascular disorder;
 KW thrombosis; wound healing.
 XX Homo sapiens.
 OS
 XX
 XX WO200155326-A2.
 XX
 XX 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001WO-US01347.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Baraeh SC, Ruben SM;
 PI
 XX WPI; 2001-451931/48.
 XX N-PSDB; AAS33264.
 DR
 XX New nucleic acids and polypeptides, useful for diagnosing, preventing
 PT or treating medical conditions -
 PT
 XX Claim 11; SEQ ID No 547; 753pp; English.
 PS
 XX The invention relates to novel isolated nucleic acid molecules (I)
 CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)

CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain tissues before transplantation, and support cell culture
 CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
 CC amino acid sequences, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 580 AA;

Query Match 99.5%; Score 2331; DB 22; Length 580;
 Best Local Similarity 99.6%; Pred. No. 8.7e-251;
 Matches 447; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTVVGSSK 60
 DB 132 MRESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTVVGSSK 191
 QY 61 FRSKERPVPVLSYLYKXENNAACRCQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVD 120
 DB 192 FRSKERPVPVLSYLYKXENNAACRCQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVD 251
 QY 121 RPKIWHFLVIMRIVLQAKNLMIDITKIFSLQDPKKEEDTDEKQALNQAQVYDND 180
 DB 252 RPKIWHFLVIMRIVLQAKNLMIDITKIFSLQDPKKEEDTDEKQALNQAQVYDND 311
 QY 181 TLDQLLRQERYKRFINSRSGWGPVGTPLRLAASYGHLSCLOVLLAHGADVDSLDVKAQTP 240
 DB 312 TLDQLLRQERYKRFINSRSGWGPVGTPLRLAASYGHLSCLOVLLAHGADVDSLDVKAQTP 371
 QY 241 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 300
 DB 372 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 431
 QY 301 KLPWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 360
 DB 432 KLPWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAAARDGAVAILQELLDHGAENVKA 491
 QY 361 CLHNCPEPIYQLIDFGANIYLPISLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 420
 DB 492 CLHNCPEPIYQLIDFGANIYLPISLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 551
 QY 421 RALCOAGQQAQINQOLDIPPMILSYLKHQL 449
 DB 552 RALCOAGQQAQINQOLDIPPMILSYLKHQL 580

RESULT 5
 AAU77411
 ID AAU77411 standard; Protein; 360 AA.
 XX
 AC AAU77411;
 XX
 XX 05-JUN-2002 (first entry)
 XX Human NOV5b protein, homologue of ankyrin-repeat proteins.
 XX Human; NOV5b-associated disorder; developmental disorder; blood disorder;
 KW endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
 KW respiratory disorder; inflammatory disorder; reproductive disorder;
 KW neurodegenerative disorder; autoimmune disorder; infectious disease;
 KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
 KW metabolic pathway modulation; NOV5b.

OS Homo sapiens.
XX WO200206329-A2.
PN 24-JAN-2002.
PD 18-JUL-2001; 2001WO-US22709.
XX 18-JUL-2000; 2000US-218870P.
XX 18-JUL-2000; 2000US-218875P.
PR 18-JUL-2000; 2000US-218901P.
PR 24-JUL-2000; 2000US-220273P.
PR 26-JUL-2000; 2000US-220912P.
PR 27-JUL-2000; 2000US-221233P.
PR 28-JUL-2000; 2000US-221650P.
XX (CURA-) CURAGEN CORP.
PA Rastelli L, Shimkets RA, Zerhusen B, Malyankar UM, Padigar M;
XX WPI: 2002-179781/23.
XX N-PSDB; ABK1107.
XX Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT polypeptides, useful for treating developmental disorders, endocrine
PT disorders, vascular disorders, infectious diseases and
PT neurodegenerative disorders -
XX Claim 1; Page 40; 178pp; English.
XX The present invention relates to the isolation of novel human
CC polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a,
CC NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC The NOVX polypeptides are related to NOPE, cadherin, interferon
CC alpha-13, ADAM, ankryn repeat-containing, transpanin or semaphorin
CC polypeptides. The sequences of the invention are useful for identifying
CC an agent (a cellular receptor or downstream effector) that binds to
CC the NOVX polypeptide, or an agent that modulates its expression or
CC activity. They are useful for treating or preventing NOVX-associated
CC disorders such as developmental disorders, endocrine disorders, vascular
CC diseases, gastrointestinal disorders, respiratory disorders, inflammatory
CC disorders, blood disorders, reproductive disorders, neurodegenerative
CC disorders, autoimmune and immune disorders, infectious diseases,
CC cardiovascular disorders, cancers, and other disorders related to cell
CC signal processing and metabolic pathway modulation. The present sequence
CC represents the human NOV5b protein.
XX Sequence 360 AA;
SQ Query Match 72.1%; Score 1690; DB 23; Length 360;
Best Local Similarity 99.4%; Pred. No. 1.7e-179;
Matches 326; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 122 PKIWHFLVLMRIVLQAKNMLDITKIFSLQPKDEEDTDEEKOALNQAVYNDST 181
DB 33 PQIWHFLVLMRIVLQAKNMLDITKIFSLQPKDEEDTDEEKOALNQAVYNDST 92
QY 182 LQQLRQERYKRPINRSRGVPGTPLRLAASVGHLSCLQVLLAHGADVSDLVKAQTEL 241
DB 93 LDQLLRQERYKRPINRSRGVPGTPLRLAASVGHLSCLQVLLAHGADVSDLVKAQTEL 152
QY 242 FTAVSHGHLDVVRVLEAGASPGSYNNCSPLVTAARGAVAILQELLDHGAENVKAK 301
DB 153 FTAVSHGHLDVVRVLEAGASPGSYNNCSPLVTAARGAVAILQELLDHGAENVKAK 212
QY 302 LPWASNIASCSGPLYLAAYVGHLDVVRVLEAGADPDYNDCTDQGLARVPRRTLEIC 361
DB 213 LPWASNIASCSGPLYLAAYVGHLDVVRVLEAGADPDYNDCTDQGLARVPRRTLEIC 272
QY 362 LHHNCEPEYIQLIDFGANIYLPFSLDLTSDQDKGIALQLQARATPRSLLSQVRLVRR 421
DB 273 LHHNCEPEYIQLIDFGANIYLPFSLDLTSDQDKGIALQLQARATPRSLLSQVRLVRR 332
SQ

QY 422 ALCQAGQPOAINQLDIPPMILSYLKHQL 449
DB 333 ALCQAGQPOAINQLDIPPMILSYLKHQL 360
RESULT 6
ABP69316
ID ABP69316 standard; Protein; 318 AA.
XX AC ABP69316;
XX 20-JAN-2003 (first entry)
DE Human polypeptide SEQ ID NO 1363.
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic.
XX Homo sapiens.
OS WO200270539-A2.
XX 12-SEP-2002.
XX 05-MAR-2002; 2002WO-US05095.
XX 05-MAR-2001; 2001US-0799451.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao OA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
DR N-PSDB; AB211533.
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders -
XX Claim 9; SEQ ID NO 1363; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (AB211119-AB212066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 318 AA;
SQ Query Match 69.8%; Score 1635; DB 23; Length 318;
Best Local Similarity 99.4%; Pred. No. 2e-173;


```
XX (HYSE-) HYSEQ INC.
XX PA
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX DR WPI; 2002-292408/33.
XX DR N-PSDB; ABN32785.
XX PT An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX PS Example 2; SEQ ID NO 867; 509pp; English.
XX CC The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention.
XX SQ Sequence 370 AA;
Query Match 27.4%; Score 642.5; DB 23; Length 370;
Best Local Similarity 53.9%; Pred. No. 2.2e-62;
Matches 145; Conservative 20; Mismatches 33; Indels 71; Gaps 8;
QY 1 MRESGKLDIPDSDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
DB 122 MRESGKLDIPDSDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 181
QY 61 FRSKERPVLVLYKNNAAICRCSOPLSGFYTRCVDDLELLEAISQTNPGSQFYVVDV 120
DB 182 FRSKERPVLVLYKNNAAICRCSOPLSGFYTRCVDDLELLEAISQTNPGSQFYVVDV 241
QY 121 RPKIWHFLVLIWRIVLQAKMNLMDITKIFSLLOPKDEEDTTEKQALNQAQVYDNDY 180
DB 242 RPK-----AVKVEKASV-----VHCSDGW 261
QY 181 -----TLDQLLRQRYKRF-----INRSRGWGPCTPLRLAASVGHLSCLQVLLAHGAD 229
DB 262 DRTAQCVSASILLDPFYTFKGLMLIEKEWISMGH--KFSQRCGHL-----D 308
QY 230 VDSLDVKAQTPFTAVSHGHLDQVRYLLE 258
DB 309 GDSKEV---SPIFTQ---FLDCIWQLME 330
RESULT 9
AAE32130
ID AAE32130 standard; Protein; 337 AA.
XX AC AAE32130;
XX DT 24-MAR-2003 (first entry)
XX DE Human cytoskeleton-associated protein, CSAP-28.
XX KW Human; cytoskeleton-associated protein; CSAP-28; atherosclerosis;
XX cancer; Gene therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 245..270
XX FT /note= "Transmembrane domain"
XX PN WO200279404-A2.
XX
```

```
PD 10-OCT-2002.
XX 25-MAR-2002; 2002WO-US09288.
XX 29-MAR-2001; 2001US-280508P.
PR 03-APR-2001; 2001US-281323P.
PR 13-APR-2001; 2001US-283769P.
PR 04-MAY-2001; 2001US-288609P.
PR 10-MAY-2001; 2001US-290518P.
PR 18-MAY-2001; 2001US-291870P.
PR 29-MAY-2001; 2001US-294451P.
XX (INCY-) INCYTE GENOMICS INC.
XX PA Hafalia AJA, Tang TY, Yue H, Khan FA, Ison CH, Baughn MR;
XX PI Warren BA, Duggan BM, Thangavelu K, Honchell CD, Azimzai Y;
XX PI Elliott VS, Burford N, Ding L, Yue H, Becha S, Emerling BM;
XX PI Richardson TW, Lee SY, Bandman O, Lal PG, Lee S, Gietzen KJ;
XX PI Wallia NK, Griffin JA, Lee EA, Swarnakar A, Ring HZ, Jones KA;
XX WPI: 2003-092894/08.
XX DR N-PSDB; AAD49617.
XX PT New human cytoskeleton-associated proteins, useful for preparing a
XX composition for diagnosing or treating a disease or condition
XX associated with decreased expression or overexpression of functional
XX CSAP e.g., cancer -
XX PS Claim 1; Page 202-203; 233pp; English.
XX CC The invention relates to new human cytoskeleton-associated protein
XX (CSAP) and its polynucleotide. The polypeptide is useful for preparing
XX a composition for diagnosing or treating a disease or condition
XX associated with decreased expression or overexpression of functional
XX CSAP e.g. atherosclerosis or cancer. The present sequence is human
XX CSAP-28 protein. The invention is useful in gene therapy.
XX SQ Sequence 337 AA;
Query Match 21.3%; Score 499; DB 24; Length 337;
Best Local Similarity 39.1%; Pred. No. 2e-46;
Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;
QY 166 EKOALNQAVYDNDSTYLDQLLRQRYKRFINSRSGW---GVPGTPLRLAASVGHLSCLQV 222
DB 39 EDRTRLDRAVYGDLTQLRLQESYRINRINRINRINRINRINRINRINRINRINRINRINR 98
QY 223 LLAHGADVDSLQVKAQTPFTAVSHGHLDQVRYLLEAGASPGSGSYNNNSPVLTAAADGA 282
DB 99 LIRKGAEDLVQKGTALYVAVVNGHLESTQILLEAGADPNRHRHRTSTPVVHASRVGR 158
QY 283 VAILQELLDHGAENNVKAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLLLH 334
DB 159 ADILKALIRYAGDVVDVNHHLTPDVQVFRFSRRLTSLVWC--PLYISAAAYHNLQCFRLLLLA 216
QY 335 GADPDYNC-----TDGGLLARVPRPTLLEICLHNCPEYIOLLIDFGANIVL---PSLS 387
DB 217 GANPDFNCGNPVNTQGFYR--GSPCGVMDAVLRHGCCEAAAFVSLVFEFGANLNLVKWESLG 274
QY 388 LDLTQS---DDKGIALLQARATPRSLLSQVRLVRRALCQAGQPOAINQLDIPPLMLISY 444
DB 275 PESRRGRKVDPEALQVFKEARSVPTLLCLCHVAVRRALGK-HRLHLIFSLPLPDPIKXF 333
QY 445 LKHQ 448
DB 334 LLHE 337
RESULT 10
AAU20362
ID AAU20362 standard; Protein; 341 AA.
XX AC AAU20362;
```


CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Caucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain organs before transplantation, and support cell culture
 CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
 CC amino acid sequences, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 343 AA;

Query Match 21.3%; Score 499; DB 22; Length 343;
 Best Local Similarity 39.1%; Pred. No. 2.1e-46;
 Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;

QY 166 EQALNQAVYDNDSTYLDOLLRQERYKRFINSRSGW---GVPTPLRLAASYGHLSCLOV 222

Db 45 ETRLHDAAYVGLQTLRLSLLQESYRSRINEKSVCCWGLPCTPLRIATAGHGCVDV 104

QY 223 LLAAGADVSLDVKAQTPFTAVSHGHLCVRLLEAGASPGGSINYNCSPVLTAARDGA 282

Db 105 LIRKGAEDLVGVKGTALYVAVVNGHLESTQILLEAGADPNRHRSTFVTHARVGR 164

QY 283 VAILQELDHGAENYKAKL-----PWASNTAS---CSGPLYLAAYVGHLCFRLLLLH 334

Db 165 ADILKALIRYGVADVDVNHHLTPDVPFRSRLTSLVVC--PLYISAAYHNLQCFRLLILA 222

QY 335 GADPDYNC---TDQGLLARVPRPTLLEICLHNCEPEYIQLIDFGANIYL---PSLS 387

Db 223 GANPDNCGPVNTQGYR--GSPGCVMDAHLRGCEAFVSLVVEFGANLNLVKWESLG 280

QY 388 LDITSQ---DDKGIALLQARATPRSLISOVRLVVRRLCAQGPQAINQLDIPPMILISY 444

Db 281 PESRGRKVDPEALQVFKEARSVPRTLCLCRVAVRRLGK-HRLHLIPLPLDPPIKPF 339

QY 445 LKHQ 448

Db 340 LLHE 343

RESULT 12

AAW62621

ID AAW62621 standard; Protein; 350 AA.

XX

AC AAW62621;

XX

DT 27-OCT-1998 (first entry)

XX Mus musculus SOCS7 protein.

XX

KW SOCS; suppressor of cytokine signalling; PCR primer;

KW autoimmune disease; diagnosis; cancer; treatment;

KW cytokine mediated cellular responsiveness; hyperimmunity;

XX immunosuppression; allergies; hypertension.

XX Mus musculus.

OS

XX Key Location/Qualifiers
 FH Misc-difference 167
 FT /note= "undefined amino acid"

XX WO9820023-A1.

XX 14-MAY-1998.

XX 31-OCT-1997; 97WO-AU00729.

XX 14-FEB-1997; 97AU-0005117.

PR 01-NOV-1996; 96AU-0003384.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Alexander WS, Hilton DJ, Metcalf D, Nicholson SE;

PI Nicola NA, Richardson RT, Starr R, Viney EM, Willson TA;

XX WPI; 1998-286854/25.

DR N-PSDB; AAV38672.

XX Suppressor of cytokine signalling proteins - useful to treat

PT disease, injury or abnormality involving cytokine mediated cellular

PT responsiveness e.g. hyperimmunity, immunosuppression, allergies and

PT hypertension

XX Claim 13; Page 144-146; 325pp; English.

XX The sequence is that of a suppressor of cytokine signalling

CC protein (SOCS). SOCS can be used to screen for naturally

CC occurring antibodies to SOCS, which may occur, e.g. in some autoimmune

CC diseases. Alternatively, specific antibodies can be used to

CC screen for SOCS, which is useful as a knowledge of SOCS levels

CC may be important for the diagnosis of certain cancers. Soluble

CC SOCS polypeptides can be used to treat disease, injury or

CC abnormality involving cytokine mediated cellular responsiveness,

CC e.g. hyperimmunity, immunosuppression, allergies and hypertension.

XX Sequence 350 AA;

Query Match 20.5%; Score 480; DB 19; Length 350;

Best Local Similarity 38.3%; Pred. No. 2.9e-44;

Matches 114; Conservative 58; Mismatches 104; Indels 22; Gaps 8;

QY 170 LNOAVYDNDSTYLDOLLRQERYKRFINSRSGW---GVPTPLRLAASYGHLSCLOVLLAH 226

Db 56 LHDAAYVGLQTLRLSLLQESYRSRINEKSVCCWGLPCTPLRIATAGHGCVDVFLRK 115

QY 227 GADVSLDVKAQTPFTAVSHGHLCVRLLEAGASPGGSINYNCSPVLTAARDGAVAIL 286

Db 116 GAEDLVGVKGTALYVAVVNGHLESTQILLEAGADPNRHRSTFVTHARVGRDDIL 175

QY 287 QELLDHGAENYKAKL-----PWASNTAS---CSGPLYLAAYVGHLCFRLLLLHGADP 338

Db 176 KALIRYGVADVDVNHHLNSDTRPPFSRLTSLVVC--PLYISAAYHNLQCFRLLLOAGANP 233

QY 339 DYNCTDQGLLARVPR--PRTLLEICLHNCEPEYIQLIDFGANIYL---PSLSLDTSQ 393

Db 234 DFNCNGPVNTQGYRSGPCVMDAHLRGCEAFVSLVVEFGANLNLVKWESLGPEARGR 293

QY 394 ---DDKGIALLQARATPRSLISOVRLVVRRLCAQGPQAINQLDIPPMILSYLKHQ 448

Db 294 RKMDPEALQVFKEARSIPRTLCLCRVAVRRLGKY-RULHLVPSPLPDPPIKFFLLYE 350

RESULT 13

AA873229

ID AAB73229 standard; Protein; 629 AA.

XX

AC AAB73229;

XX

DT 11-MAY-2001 (first entry)

XX Human phosphatase MTR7 h.
DE Phosphatase; gene therapy; enzyme; cancer; pathophysiological hypoxia;
XX cardiac dysfunction; vascular disorder; myopathy; ectodermal dysplasia;
KW congenital muscle disorder; Papillon-Lefevre syndrome; Cowden disease;
KW Moebius syndrome; Bjornstad syndrome; Bannayan Zonana syndrome;
KW schizoprenia; hamartoma.
XX Homo sapiens.
OS WO200112819-A2.
XX 22-FEB-2001.
XX 11-AUG-2000; 2000WO-US22158.
XX 13-AUG-1999; 99US-0149005.
XX (SUGB-) SUGEN INC.
XX Plowman GD, Martinez R, Whyte D, Hill R, Flanagan P, Lioubin M;
PI WOPI; 2001-211226/21.
DR N-PSDB; AAF63581.
XX New protein phosphatase polypeptide for diagnosing and treating
PT phosphatase related disorders such as cancer, schizoprenia, cardiac
PT dysfunction and/or vascular disorders
XX Claim 6; Fig 5; 138pp; English.
XX The present invention relates to phosphatase proteins and coding
CC sequences. The present sequence is one such phosphatase. Phosphatases are
CC enzymes that catalyse the dephosphorylation of proteins modified by
CC phosphorylation of serine, threonine or tyrosine residues. The
CC phosphatases are useful for treating a variety of diseases: for example
CC cancer e.g. breast, urogenital, prostate, head, neck, lung cancers,
CC synovial sarcomas, renal cell carcinoma, non-small cell lung cancer,
CC hepatocellular carcinoma, pancreatic endocrine tumours, stomach cancer,
CC glioblastoma, colorectal cancer and thyroid cancer, pathophysiological
CC hypoxia, cardiac dysfunction and/or vascular disorders, myopathies,
CC congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease,
CC ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan
CC Zonana syndrome, schizoprenia and hamartomas.
XX Sequence 629 AA;
SQ Query Match 18.1%; Score 423; DB 22; Length 629;
Best Local Similarity 60.2%; Pred. No. 1.8e-37;
Matches 74; Conservative 21; Mismatches 28; Indels 0; Gaps 0;
QY 2 RESGWLIDPISDFGRMGIPNRNWTITDANRNYEISTYPPETVVPKSVTLGTVVGGSKF 61
DB 197 REQQWLIDLSEBYTRMGLPNHYWQLSDVNRDVRVCDSTYPTLYPKSATAHIIVGSSKF 256
QY 62 RSKERVPVLSYLYKNNAAICRCSOPLSGFYTRCVDDLELLAISQTNPGSQFMYVVDTR 121
DB 257 RSRREFVLSYLYKNDHASICRSSOPLSGFSARCLEQMLQAIRKANPGSDFVYVVDTR 316
QY 122 PKI 124
DB 317 PKL 319
RESULT 14
ID ABP62857
XX ABP62857 standard; Protein; 482 AA.
AC ABP62857;
XX 14-OCT-2002 (first entry)
DT
XX

DE Human polypeptide SEQ ID NO 294.
XX Human; vulnary; dermatological; neuroprotective; nootropic; cancer;
KW antiparkinsonian; immunostimulant; cytostatic; immunosuppressive;
KW antidiabetic; antiallergic; gene therapy; wound healing; tissue repair;
KW burn; central nervous system disorder; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; immune disorder;
KW autoimmune disease; multiple sclerosis; diabetes; allergy.
XX Homo sapiens.
OS WO200218424-A2.
XX 07-MAR-2002.
XX 31-AUG-2001; 2001WO-US27093.
XX 01-SEP-2000; 2000US-0654935.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
PI Zhao QA, Wang D, Liu C, Drmanac RT, Wehrman T;
XX WPI; 2002-583321/62.
DR N-PSDB; ABQ93336.
XX New polynucleotide and polypeptides, useful for treatment and diagnosis
PT of Alzheimer's, Parkinson's, Huntington's, amyotrophic lateral
PT sclerosis, immune deficiencies, cancer, autoimmune disorders, multiple
PT sclerosis, diabetes and allergies
XX Claim 20; SEQ ID NO 294; 284pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising one of
CC 245 sequences (ABQ93288-ABQ93532). Treating a condition comprising
CC administering to a mammalian subject a composition comprising the protein
CC (II) encoded by (I) (ABP62809-ABP63053) or an antibody (III) to (II).
CC (I), (II) and (III) are useful for diagnostic evaluation of disorders.
CC (I) is useful for gene therapy of diseases and (II) can be used for
CC therapeutic treatment. Diseases that may be treated include wound healing
CC and tissue repair, burns, central nervous system disorders (e.g.
CC Alzheimer's, Parkinson's, Huntington's and amyotrophic lateral
CC sclerosis), immune deficiencies, cancer, autoimmune disorders, multiple
CC sclerosis, diabetes and allergies
CC Note: the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 482 AA;
SQ Query Match 16.8%; Score 392.5; DB 23; Length 482;
Best Local Similarity 44.6%; Pred. No. 2.9e-34;
Matches 74; Conservative 22; Mismatches 27; Indels 43; Gaps 1;
QY 2 RESGWLIDPISDFGRMGIPNRNWTITDANRNY----- 34
DB 7 REQQWLIDLSEBYTRMGLPNHYWQLSDVNRDVRVCDSTYPTLYPKSATAHIIVGSSKFGIITDFRGLD 66
QY 35 -----EICSTYPPETVVPKSVTLGTVVGGSKFRSKERVPLVLSYLYKNN 78
DB 67 FSCGFSTPTTSSRVIKVCDSTYPTLYPKSATAHIIVGSSKFRSRRRFPVLSYLYKDNH 126
QY 79 AAICRCSOPLSGFYTRCVDDLELLAISQTNPGSQFMYVVDTRPKI 124
DB 127 ASICRSSOPLSGFSARCLEQMLQAIRKANPGSDFVYVVDTRPKL 172
RESULT 15
ID ABB65465
XX ABB65465 standard; Protein; 639 AA.
AC ABB65465;

XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 23187.
DE Drosophila melanogaster polypeptide SEQ ID NO 23187.
XX Drosophila melanogaster polypeptide SEQ ID NO 23187.
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL09568.
DR N-PSDB; ABL09568.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 23187; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 639 AA;
SQ
Query Match 15.4%; Score 360; DB 22; Length 639;
Best Local Similarity 26.9%; Pred. No. 2e-30;
Matches 119; Conservative 62; Mismatches 159; Indels 102; Gaps 15;
QY 2 RESGKLIIDPISDFGRMGIPNRRWTITDANRNYEIGCTYPPEIVVPKSVTLGTVVGSSKF 61
DB 71 KNAGWDYFKLEAEFKMLVNEAWTLCWNEKYELCDTPRQIYVPKREATTMLISSRF 130
QY 62 RSKERVPLSYLKENNAACRCQPLSGFYTRCVDDELLEAISQTFNPGSQPMYVVDTR 121
DB 131 RSKGRPLVLYLH-NNKASICRCQPLSGFSARCLEDQMLEAIRKTNSTNDYMYVVDTR 189
QY 122 PKI-----WHFLVLMRIVLQAKMMLMDITKIFSLLPDKEE 159
DB 190 PRINAMANAAGKGYENAEFYENIKFHLG-ENIHVQRASLQ-----KVLEACEQK 240
QY 160 EDTDTEKQALN-----QAVYDNDSTYLDLQERYKRFINSRSGWVPG-----T 206
DB 241 SPTMSAFINALESSGWLKHSIRLSTSSFTANAV--DKGVSVVYVHCSDGMDRTAQVCSLA 298
QY 207 PLFLAASYGHLSCLOVL-----LAHGADV-----SLDVKAOTPLFTAVSHGLDCV-- 253
DB 299 QLMNLPYRTIKGFQALIEKDWLAFGHKFSERCCHIQTDAREVSPFTQ-----FLDCTWQ 354
QY 254 -----RVLL-----EAGSPGGSIVNCSFVLTAARDGAVAILQELLHDGAEA 296
DB 355 LMSQRSEAFEFNERFLILHDVHVSQCFGTGVCNCE-----KORLDLKLAE 400

QY 297 NVKAKLPVWASNIASCSGPLYL-----AAVYGHLDGCFRLLLLHGADPDYNCYDQGLIARVP 352
DB 401 RTFSLWGYMANHLNEYINPLYKPNVDEAIKANLAPQCICKFWEGM---YSRPFSGIHPREP 457
QY 353 RPRTLLEICLHNCBPEYIQLL 374
DB 458 LGDVLDSKEHCNLSLEDHVQHL 479

Search completed: December 1, 2003, 07:37:11
Job time : 72 secs

THIS PAGE BLANK (USPTO)

```
RESULT 5
US-09-155-078-2
; Sequence 2, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra
; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078
; CURRENT FILING DATE: 1999-03-02
; EARLIER APPLICATION NUMBER: PCT/DE97/00592
; EARLIER FILING DATE: 1997-03-21
; EARLIER APPLICATION NUMBER: DE 196 11 234.6
; EARLIER FILING DATE: 1996-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 621
; TYPE: PRT
; ORGANISM: German-type microorganism & cell cul.
US-09-155-078-2

Query Match
Best Local Similarity 92.7% Score 264; DB 4; Length 621;
Matches 80; Conservative 46; Mismatches 106; Indels 58; Gaps 12;

QY 5 GWKLDIPIDFGRMGIPNRNMTITDANRYETCSYTPPEIVVPKSVTIGTVVGGSKFRSK 64
DB 181 GWTVPVVEYRROGLPNHNRITFTKCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240
QY 65 ERVPVLSVLYKNNAAICRCSQPLSGFY-TRCVDELLLEALSQNPFGSQFYVVDTRPK 123
DB 241 NRIPVLSWIHPENKTIVVRCQPLVMSGKRNKDEKYLVDVIRETNKQISKLTIIYDARPS 300
QY 124 I-----WHFLVLMRIVLQAKNLM--DITKIFSLQDPKBE-----159
DB 301 VNAVANKATGGYESSDDAYHNAELFF---LDHNIHVRESLKKVKDVIYFNVESHWS 357
QY 160 --EDTTEE--KQALNQAVYDNDSTYLDQLQRYKRFINSRSGVPGTPLRLA----211
DB 358 SLESTHWHLEHKLVLTAIQVADK-----VSSGKSSVLVHSCDGDWDRTAQLTSLAML 411
QY 212 -----ASYGHLSCLOVLLAHGADVSLDVKATQPLFTAVSHGHLCV 253
DB 412 DSFYRSIEGFELVQKEWISFGHKFASRI--GHG-DKNHTDAD-RSPIFLQ----FIDCV 463

RESULT 6
US-09-350-982C-5
; Sequence 5, Application US/09350982C
; Patent No. 6455290
; GENERAL INFORMATION:
; APPLICANT: Berthelsen, Jens
; APPLICANT: Toma, Salvatore
; APPLICANT: Isacchi, Antonella
; TITLE OF INVENTION: Tankyrase Homolog Protein(THP), Nucleic Acids, and Methods Rela
; TITLE OF INVENTION: Same
; FILE REFERENCE: PHRM-0043
; CURRENT APPLICATION NUMBER: US/09/350,982C
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 5
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Artificial
```

```
FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: (1102)..(1102)
; OTHER INFORMATION: n is any nucleic acid
; NAME/KEY: misc feature
; LOCATION: (2650)..(2650)
; OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-5

Query Match
Best Local Similarity 9.3% Score 219; DB 4; Length 1166;
Matches 74; Conservative 26; Mismatches 78; Indels 66; Gaps 7;

QY 189 ERVKRF-----INRSRGVPGTPLRLAASGYHLSCLQVLLAHGADVSLDVKATQPLFT 243
DB 38 ERVKRLVTPEKVNSTRDTAGRKSTPLHLAAGFGKVDVVEYLLONGANYOARDGGLIPLHN 97
QY 244 AVSHGHLDVVRVLEAGASPGSGSYNNCSPLVLTAAARDGAVAILQELLDHGAANVKAKLP 303
DB 98 ACSFGHAEVNVNLLRHGADPNARDNWNYPPLHEAAIKGKIDVCIVLLQHGAEPTIRNTDG 157
QY 304 VWASNIA-----SC-----SGPYL 318
DB 158 RTALDLADPSAKAVLTGCKYKDELLESARSNGEKKMALLTPLNVNCHASDGRKSTPLHL 217
QY 319 AAVYGHLDLFRLLHGHADPNCTDQGLLARVPRPTLLEICLHNC---EPEYIQLLI 375
DB 218 AAGYNRYKIVQLLQHGAA--DVHAKRKGDL--VP-----LHNACSYGHYEVTLLV 264
QY 376 DFGA 379
DB 265 KHGA 268

RESULT 7
US-09-196-387-10
; Sequence 10, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
```

THIS PAGE BLANK (USPTO)

Db 347 GGGGCCGAGGTGGACCTGGTGGATGTCAGAGGGCAGACTGCCCTGTATGTGGCTGTAGTG 406
Qy 1105 CATGGCCATCTGACTGTGTACGTGTCTTTTGGAGCTGGTGGCTCTCTCTGTGTGTAGC 1164
Db 407 AACGGGCACCTGGAGAGCACTGAGATCTTTTGGAGCTGGTGTATGCCCAACGGCAGC 466
Qy 1165 ATCTACAAACAACTGTTCTCCCGTCTCACAGCTGCCCGTATGCTGTGTGTCTATCTCTG 1224
Db 467 CGGCACACCGCAGCATCTCTGTATACATGCCCTTCTGTGTGGTAGGAGCAGCATCTCTG 526
Qy 1225 CAGAGCTCTAGACCATGTGTGAGGGCAGGCAAGCTCAAGCTAACT-----ACC 1274
Db 527 AAGGCTTATCAGGTATGGGCAGATGTTGATGTCAACCATCATCTGAATTTCTGACACC 586
Qy 1275 AGTCTGGGCATCAACATAGCTTCTCT-----GGCCCCCTCTATTTGGCGCA 1326
Db 587 CGGCCCCCTTTTTCAGGGGCTAACCTCTTGTGTGTCTCTCTATACATCAGTGTCT 646
Qy 1327 GTCTACGGGCACCTGGAGCTGTTTCCGCTCTTGTCTTCCAGGGGCGAGACCCTGACTAC 1386
Db 647 GCCTACCATCACTTCACTGCTTTCAGGCTGCTTTCAGGCTGGGGCAATCTCTGACTTC 706
Qy 1387 AACTGCACTGACAGGGC 1404
Db 707 AATTGCAATGGCCCTGTC 724

RESULT 2

US-09-302-769-26
; Sequence 26, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; PRIOR FILING DATE: 1999-04-30
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 26
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Human
; US-09-302-769-26

Query Match 3.8%; Score 71.2; DB 4; Length 419;
Best Local Similarity 55.4%; Pred. No. 2.2e-11;
Matches 164; Conservative 0; Mismatches 123; Indels 9; Gaps 1;
Qy 862 GAGAGCAGGCTCTCAATCAAGCAGTGTATGACAAAGCTCTCTATCTTTGGACGACTT 921
Db 114 GAGACACAGGCTCCATGATGACGCTTACGTGGGAGCTCCAGACCTCAGAGGCTTA 173
Qy 922 TTGGCCAGGAGCGTTTCAACGTTTCAACAGCAGGAGTGGCTGG-----GGT 972
Db 174 TTGAAGAGGAGAGCTACCGGAGCGCATCAACAGAGTCTGTCTGTGTGGCTGG 233
Qy 973 GTTCTGGGACACCTTGGCTGCTTCTTATGGCCACTGAGCTGTTTCCAGTTC 1032
Db 234 CTCCTCTGCACACCGTTTGCAGTTCGGGCACTCGAGGCCATGGGAGCTGTGTGACTTC 293
1033 CTCCTAGCCCATGCTGTGTATGTTGACAGCTTGGATGTCAGGCGCACAGCCCACTTTTC 1092

Db 294 CTCATCCGGAAGGGCGGAGGTGGATCTGTGGAGCTGTAAGAGACAGACGGCCCTGTAT 353
Qy 1093 ACTGCTGTGACGTCATGGCCATCTGACTGTGTACGTGTCTTTTGGAGCTGGTGC 1148
Db 354 GTGGCTGTGTGAACCGGCACCTAGAGAGTACCCAGATCTCTCTCGAAGCTGGCGC 409
RESULT 3
US-09-221-298-28
; Sequence 28, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: XU, Jianshun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: modified_base
; FEATURE:
; LOCATION: (212)
; OTHER INFORMATION: Where n is a, c, g or t
; US-09-221-298-28

Query Match 3.7%; Score 69.4; DB 3; Length 401;
Best Local Similarity 52.4%; Pred. No. 7.4e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
Qy 360 AAAAGAGATGAGGGAAGTGGATGGAACCTGATTTGACCAATATCAGACTTTGGGCGTAT 419
Db 52 AGAAAAGTTTAACTGGATGGATGGACAGTTTACAAATCCAGTGAAGAATACAGGAGGCA 111
Qy 420 GGGATACCCACAGAACTGGACCAATACAGATGCCAAGAACTATGAGATATGCAG 479
Db 112 GGGCTTGGCCCAATCACCATTGGAGATAACTTTTATTAATAAGTGTCTATGAGCTCTGCGA 171
Qy 480 CACTACCTCTCTGAAATAGTGTCTTAACTCTTACCTTGGGAACGGTGTGGAAG 539
Db 172 CACTACCTCTCTTGTGTGTCTTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 231
Qy 540 TTCAAAGTTTCAAGATTAAGAAAGCTGCTCCCTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 599
Db 232 TGCAACTTTTAGTCCCGAAATCGAATTCAGTGTGTCTATGAGTTCATCCAGAAATAA 291
Qy 600 TGCTGCCATTTGGCGCTGTAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
Db 292 GACGGTCATTGTGCGTTGCACTGAGCTCAGCTCTTGTGCGTATGAGTGGGAAACGAAATAAGA 351
Qy 657 TGATGAGCTCTTGTGGAGGCCATTAGCCAAACAAA 692
Db 352 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 387

RESULT 4

US-09-155-078-1
; Sequence 1, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra
; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078

CURRENT FILING DATE: 1999-03-02
EARLIER APPLICATION NUMBER: PCT/DE97/00592
EARLIER FILING DATE: 1997-03-21
EARLIER APPLICATION NUMBER: DE 196 11 234.6
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 3431
TYPE: DNA
ORGANISM: German-type microorganism & cell cul.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1863)
US-09-155-078-1

Query Match 3.6%; Score 67.2; DB 4; Length 3431;
Best Local Similarity 52.1%; Pred. No. 1.3e-09;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

QY 360 AAAAGATGAGGAGGAAAGTGTGAACTGATGACCCCAATTCAGACTTTGGCGTAT 419
DB 522 AAAAAAGTTAACTGTGATGATGAGACAGTTTACATTCAGTGAAGATACAGAGGCA 581
QY 420 GGAATACCCACAGAACTGACCATACAGATGACCAAGAACTATGATATGACG 479
DB 582 GGGCTTGGCCCAATCACCATTGAGAAATCTTTATTAATAGTGTATGAGCTGTGA 641
QY 480 CACTTACCTCTGTAATAGTGTCTTAATCTGTTACCTTGGAAACGGTGTGAG 539
DB 642 CACTTACCTCTGTTGT 701
QY 540 TTCAAAAGTTCAAGATGAAAGAGTGTCTCTGTCTCTCACTCTCAAGAGACAA 599
DB 702 TCCAACTTTAGGTCCCGAAATCGAATTCAGTGTGTATGATTCATCCAGAAATTA 761
QY 600 TGTCTGCAATTTGCGGTGAGCAAGCTCTCTGTGATTTTACACTGCGTGTAGA 656
DB 762 GAGGCTCATGTGCGGTGAGTCAAGTCTCTGTGTGTGTGTGTGTGTGTGTGTGT 821
QY 657 TGATGAGCTCTGTGAGAGGCAATGACCAACAA 692
DB 822 TGATGAGAAATATCTGATGTTATCGAGAGACTTA 857

RESULT 5

US-08-181-629A-2
Sequence 2, Application US/08181629A
Patent No. 5472872

GENERAL INFORMATION:
APPLICANT: Swaminathan, Neela
APPLICANT: Van Elten, James
APPLICANT: Mead, David
APPLICANT: Skowron, Piotr
TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,629A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31504
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-0448
TELEFAX: 312/474-6300
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5496 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-181-629A-2

Query Match 3.5%; Score 66.6; DB 1; Length 5496;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 144; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 978 TGGACACCTTGGCGCTGCTCTTTATAGCCATTGAGCTGTTGCAAGCTCTT 1037
DB 3468 TTGACGGCGTTACATTAAGCGGCTTTAATGTCATAGCATGTCGCAAGACCTTAT 3527
QY 1038 AGCCATGCTGTATGTTGACAGTGTGATGTCAAGCAGAGCCACTTTTCACTGC 1097
DB 3528 TGATCGGGGTGCAATCTTGACATCAGATATTTGGAGTGTACACACTTCATGTCG 3587
QY 1098 TGTCAATGAGCCATCTGACATGTCATGTCGTCTTTGAGAGTGTGCTCTCG 1157
DB 3588 GGTATATATACACAGATGATGTATGATTAAGTCTGTAGAGAGTCAACTCTTGA 3647
QY 1158 TGTAGATCTTACAAACATGTTCTCCGCTCTACAGCTGCGGTGATGCTGTTC 1217
DB 3648 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3707
QY 1218 TATCTGAGAGAGCTCTTCAAGCATGTGTCAGA 1250
DB 3708 GATTTTGAAGATGCTCATTAAGAGAGGTGAGA 3740

RESULT 6

US-09-509-802-1
Sequence 1, Application US/09509802
Patent No. 6489130

GENERAL INFORMATION:
APPLICANT: Immunex Corp.
APPLICANT: Bird, Timothy
APPLICANT: Vitca, G.D.
TITLE OF INVENTION: DEATH ASSOCIATED KINASE CONTAINING ANKYRIN REPEATS (DAKAR)
FILE REFERENCE: 2889-US
CURRENT APPLICATION NUMBER: US/09/509,802
CURRENT FILING DATE: 2000-06-02
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent In version 3.0
SEQ ID NO 1
LENGTH: 2370
TYPE: DNA
ORGANISM: Mus sp.
US-09-509-802-1

Query Match 3.2%; Score 60.2; DB 4; Length 2370;
Best Local Similarity 51.3%; Pred. No. 1.2e-07;
Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 980 GACACCTTGGCGCTGCTCTTTATAGCCATTGAGCTGTTGCAAGCTCTTATG 1039
DB 1829 GACACCTTGGCGCTGCTCTTCAAGAGGGGCAATTAACGTTGCTGCAATTCATTTG 1888
QY 1040 CCGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1099
DB 1889 ACTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1948

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:37:16 ; Search time 539 Seconds
(without alignments)
153.646 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESCWKLIPTSDPGRMGI.....QAINQLDIPPLMLSYLKHQL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2343	100.0	449	10	US-09-941-831-20
2	2343	100.0	449	12	US-10-207-175-18
3	1690	72.1	360	10	US-09-908-193-14
4	1629	69.5	351	10	US-09-908-193-12
5	1435	61.2	308	10	US-09-908-193-36
6	499	21.3	335	10	US-09-908-193-37
7	481	20.5	336	10	US-09-908-193-38
8	480.5	20.5	327	10	US-09-908-193-39
9	480	20.5	350	10	US-09-908-805B-25
10	454	19.4	271	10	US-09-908-193-40
11	233	12.5	458	11	US-09-811-469-10
12	293	12.5	458	11	US-09-811-469-11
13	293	12.5	458	12	US-10-370-659-10
14	293	12.5	458	12	US-10-370-659-11
15	264	11.3	603	11	US-09-919-039-134

16	246	10.5	323	9	US-09-884-870-2	Sequence 2, Appli
17	223	9.5	1181	12	US-10-199-937-139	Sequence 139, App
18	217.5	9.3	1262	12	US-10-199-937-107	Sequence 107, App
19	217.5	9.3	1385	12	US-10-199-937-133	Sequence 133, App
20	215	9.2	522	12	US-10-199-937-99	Sequence 99, Appli
21	215	9.2	1074	9	US-09-509-196A-2	Sequence 2, Appli
22	215	9.2	1166	11	US-09-972-115A-6	Sequence 6, Appli
23	215	9.2	1166	12	US-10-199-937-135	Sequence 135, App
24	215	9.2	1166	15	US-10-163-587A-15	Sequence 15, Appli
25	215	9.2	1169	12	US-10-199-937-2	Sequence 2, Appli
26	215	9.2	1169	12	US-10-199-937-101	Sequence 101, App
27	215	9.2	1227	12	US-09-849-602-26	Sequence 26, Appli
28	214	9.1	756	12	US-10-199-937-91	Sequence 91, Appli
29	214	9.1	784	12	US-10-199-937-89	Sequence 89, Appli
30	209	8.9	949	9	US-09-841-835-10	Sequence 10, Appli
31	208.5	8.9	1327	9	US-09-841-835-2	Sequence 2, Appli
32	208.5	8.9	1327	11	US-09-972-115A-8	Sequence 8, Appli
33	208.5	8.9	1327	12	US-10-199-937-4	Sequence 43, Appli
34	205	8.7	1724	10	US-09-964-899-43	Sequence 117, App
35	203	8.7	1762	12	US-10-205-194-117	Sequence 3, Appli
36	197	8.4	1184	12	US-10-117-229-3	Sequence 9, Appli
37	197	8.4	1184	12	US-10-117-229-9	Sequence 4, Appli
38	197	8.4	1771	12	US-10-117-229-4	Sequence 11, Appli
39	197	8.4	1771	12	US-10-117-229-11	Sequence 4, Appli
40	196.5	8.4	1715	12	US-10-021-571-4	Sequence 2, Appli
41	196.5	8.4	1715	12	US-10-117-229-2	Sequence 4, Appli
42	195	8.3	1619	12	US-10-369-978-4	Sequence 2, Appli
43	195	8.3	1704	12	US-10-369-978-2	Sequence 41, Appli
44	194.5	8.3	802	10	US-09-964-899-41	Sequence 7, Appli
45	194	8.3	1762	12	US-10-117-229-7	

ALIGNMENTS

RESULT 1

US-09-941-831-20
; Sequence 20, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: PT049P1
; FILE REFERENCE: Serine/Threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941,831
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-831-20

Query Match 100.0%; Score 2343; DB 10; Length 449;
Best Local Similarity 100.0%; Pred. No. 4.2e-233;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRESCWKLIPTSDPGRMGIPTNRNWTITDANRNYEICSTYPPETVVPKSVTLGTVVGSSK	60
Db	1	MRESCWKLIPTSDPGRMGIPTNRNWTITDANRNYEICSTYPPETVVPKSVTLGTVVGSSK	60
Qy	61	FRSKERVPLSVLYKENNAACRCQSPISGFTYRCVDDLELLAISQTNPGSQFMVVDVT	120
Db	61	FRSKERVPLSVLYKENNAACRCQSPISGFTYRCVDDLELLAISQTNPGSQFMVVDVT	120
Qy	121	RPKIWHFLVIMRIVLQAKNMLMDITKIFSLQPKDEEDTDTTEKQALNQAQVYDNDSY	180
Db	121	RPKIWHFLVIMRIVLQAKNMLMDITKIFSLQPKDEEDTDTTEKQALNQAQVYDNDSY	180

Db 273 LHHNCEPEYIQLLIDFGANIYPSLSLSDTSQDDKGIALLLQARATPRSLLSQVRLVVR 332

QY 422 ALCAQOQOAINOLDIPPMILSYLKHQL 449
|||||

Db 333 ALCAQOQOAINOLDIPPMILSYLKHQL 360
|||||

RESULT 4
US-09-908-193-12
; Sequence 12, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX polypeptide
; OTHER INFORMATION: with homology to ankryrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-12

Query Match 69.5%; Score 1629; DB 10; Length 351;
Best Local Similarity 97.5%; Pred. No. 1.8e-159;
Matches 318; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 132 MRIVLQAKNLMIDITKIFSLQPDKEEDTDTEEKQALNQAVYDNDSTYLLDQLLRQERY 191
|||||

Db 26 MRIVLQAKNLMIDITKIFSLQPDKEEDTDTEEKQALNQAVYDNDSTYLLDQLLRQERY 85
|||||

QY 192 KRFNRSRGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPLFTAVSHGLD 251
|||||

Db 86 KRFNRSRGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPLFTAVSHGLD 145
|||||

QY 252 CVRVLLLEAGSPGSIYNNCSPLVLTAAARDGAVAILQELLDHGAENAVKAKLPWASNIAS 311
|||||

Db 146 CVRVLLLEAGSPGSIYNNCSPLVLTAAARDGAVAILQELLDHGAENAVKAKLPWASNIAS 205
|||||

QY 312 CSGPLYLAAYVGHLDLFRLLLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYI 371
|||||

Db 206 CSGPLYLAAYVGHLDLFRLLLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYI 265
|||||

QY 372 QLLIDFGANIYPSLSLSDTSQDDKGIALLLQARATPRSLLSQVRLVVRAL 423
|||||

Db 266 QLLIDFGANIYPSLSLSDTSQDDKGIALLLQARATPRSLLSQVRLVVRAL 325
|||||

QY 424 COAQOQOAINOLDIPPMILSYLKHQL 449
|||||

Db 326 COAQOQOAINOLDIPPMILSYLKHQL 351

RESULT 5
US-09-908-193-36
; Sequence 36, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-908-193-36

Query Match 61.2%; Score 1435; DB 10; Length 308;
Best Local Similarity 88.6%; Pred. No. 1.7e-139;
Matches 273; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

QY 141 NMLMDITKIFSLQPDKEEDTDTEEKQALNQAVYDNDSTYLLDQLLRQERYKRFNRSRG 200
|||||

Db 1 NMLMDITKIFSLQPDKEEDTDTEEKQALNQAVYDNDSTYLLDQLLRQERYKRFNRSRG 60
|||||

QY 201 WGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPLFTAVSHGLDCVRVLEAG 260
|||||

Db 61 WGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPLFTAVSHGLDCVRVLEAG 120
|||||

QY 261 ASPGSIYNNCSPLVLTAAARDGAVAILQELLDHGAENAVKAKLPWASNIASCSGPLYLAA 320
|||||

Db 121 ASPGSIYNNCSPLVLTASRDGAFILQELLHGAEANVAKAKLPWASNIASCSGPLYLAA 180
|||||

QY 321 VYGHLDLFRLLLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYIQLLIDFGAN 380
|||||

Db 181 VYGHLDLFRLLLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYIQLLIDFGAN 240
|||||

QY 381 IYLPISLSDTSQDDKGIALLLQARATPRSLLSQVRLVVRALCOAGOPQAINOLDIPPM 440
|||||

Db 241 IYLPISLSDTSQDDKGIALLLQARATPRSLLSQVRLVVRALCOAGOPQAINOLDIPPM 300
|||||

QY 441 LISYLKHQ 448
|||||

Db 301 LISYLKHQ 308
|||||

RESULT 6
US-09-908-193-37
; Sequence 37, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:

APPLICANT: RASTELLI, LUCA
APPLICANT: SHIMKETS, RICHARD A.
APPLICANT: ZERHUSEN, BRYAN
APPLICANT: MALYANKAR, URIEL M.
APPLICANT: PADIGARU, MURALIDHARA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 21402-062
CURRENT APPLICATION NUMBER: US/09/908,193
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: 60/220,273
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/221,650
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/221,233
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/220,912
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/218,875
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/218,875
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/218,870
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,901
PRIOR FILING DATE: 2000-07-18
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 37
LENGTH: 335
TYPE: PRT
ORGANISM: Homo sapiens
US-09-908-193-37

Query Match 21.3%; Score 499; DB 10; Length 335;
Best Local Similarity 39.1%; Pred. No. 1.1e-42;
Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;
QY 166 EQALQAAVYDNDSTYLDQLLRQERYKRFNRSRW---GVGPTPLRLAASVGHLSCLQV 222
DB 37 EDRLHDAAYVGDQLRLSLLQESYRSRINEKSVCCGWLPTCTPLRIATAAGHSCVDF 96
QY 223 LLAHAGADVSLDKAQTPLPTAVSHGHLCVRLLEAGASPGSIVNNCSVPLTAARDGA 282
DB 97 LIRKGAEDLVKGTALYVAVVNGHLESTQILLEAGADPNRSRHRSTPVVHARSVGR 156
QY 283 VATLQELDHGAANYKAKL-----PWASNIAS---CSGPLYLAAYVGHLCDFRLLH 334
DB 157 ADTLKALIRYADVNVNHLTPDVQPRFSRRLTSLVVC--PLVISAAYHNLCQFRLLLA 214
QY 335 GADPDYNC---TDQGLLARVPRPTLLEICLHNCEPEYIQLLIDFGANIYL---PSLS 387
DB 215 GANPDNCGPVNTQGFYR--GSFGCVMDAVLRHGCEAAVSVLLVFEFGANLVKWSLG 272
QY 388 LDLSQ---DDKGIALLLQARATPRSLLSQVRLVRRALCOAQOQAINOLDIPPMILSY 444
DB 273 PESRGRKVDPEALQVFEKARSVPTLLCLCRVAVRRALGK-HRLHLPSLPDPPIKKF 331
QY 445 LKHQ 448
DB 332 LKHE 335

RESULT 7
US-09-908-193-38
Sequence 38, Application US/09908193
Publication No. US20020192748A1
GENERAL INFORMATION:
APPLICANT: RASTELLI, LUCA
APPLICANT: SHIMKETS, RICHARD A.
APPLICANT: ZERHUSEN, BRYAN
APPLICANT: MALYANKAR, URIEL M.
APPLICANT: PADIGARU, MURALIDHARA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 21402-062
CURRENT APPLICATION NUMBER: US/09/908,193

CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: 60/220,273
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/221,650
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/221,233
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/220,912
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/218,875
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,870
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,901
PRIOR FILING DATE: 2000-07-18
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 336
TYPE: PRT
ORGANISM: Mus musculus
US-09-908-193-38

Query Match 20.5%; Score 481; DB 10; Length 336;
Best Local Similarity 38.3%; Pred. No. 7.9e-41;
Matches 114; Conservative 59; Mismatches 103; Indels 22; Gaps 8;
QY 170 LNAQVYDNDSTYLDQLLRQERYKRFNRSRW---GVGPTPLRLAASVGHLSCLQVLLAH 226
DB 42 LHDAAYVGDQLRLNLLQESYRSRINEKSVCCGWLPTCTPLRIATAAGHNCVDFLIRK 101
QY 227 GADVSLDKAQTPLPTAVSHGHLCVRLLEAGASPGSIVNNCSVPLTAARDGAVATL 286
DB 102 GAEVLDVVKGTALYVAVVNGHLESTQILLEAGADPNRSRHRSTPVVHARSVGRDIL 161
QY 287 QELLDHGAANYKAKL-----PWASNIAS---CSGPLYLAAYVGHLCDFRLLHGDGP 338
DB 162 KALIRYADVNVNHLTPDTPRPFSSRRLTSLVVC--PLVISAAYHNLCQFRLLQAGNP 219
QY 339 DYNDTQGLLARVPR--PRTLLEICLHNCEPEYIQLLIDFGANIYL---PSLSLDLSQ 393
DB 220 DFNCGPVNTQGFYRSGPCVMDAVLRHGCEAAVSVLLVFEFGANLVKWSLGPEARGR 279
QY 394 ---DDKGIALLLQARATPRSLLSQVRLVRRALCOAQOQAINOLDIPPMILSYLKHQ 448
DB 280 RKMDPEALQVFEKARSIPRTLILSLCRVAVRRALGKY-RHLVPSLPDPPIKKFLIYE 336

RESULT 8
US-09-908-193-39
Sequence 39, Application US/09908193
Publication No. US20020192748A1
GENERAL INFORMATION:
APPLICANT: RASTELLI, LUCA
APPLICANT: SHIMKETS, RICHARD A.
APPLICANT: ZERHUSEN, BRYAN
APPLICANT: MALYANKAR, URIEL M.
APPLICANT: PADIGARU, MURALIDHARA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 21402-062
CURRENT APPLICATION NUMBER: US/09/908,193
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: 60/220,273
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/221,650
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/221,233
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/220,912
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/218,875
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,870

; PRIOR FILING DATE: 2000-07-18
 ; PRIOR APPLICATION NUMBER: 60/218,901
 ; PRIOR FILING DATE: 2000-07-18
 ; NUMBER OF SEQ ID NOS: 74
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 39
 ; LENGTH: 327
 ; TYPE: PR1
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: MOD RES
 ; LOCATION: (314)
 ; OTHER INFORMATION: Any Amino Acid
 ;
 ; US-09-908-193-39

 Query Match 20.5%; Score 480.5; DB 10; Length 327;
 Best Local Similarity 40.7%; Pred. No. 8.5e-41;
 Matches 111; Conservative 51; Mismatches 90; Indels 21; Gaps 7;

 QY 170 LNOAVYDNDSYTLQDLQERKFRFNSRSGW---GVPGTPLRLAASVGHLSCLQVLLAH 226
 DB 34 LHDAAYVGDQLTLRLNLQESYRSRINEKSVWCCGMLPCTPLRIAATAAGHCNCVDFLRK 93

 QY 227 GADVSLDVKAQOTPLFTAVSHGHLDVVRVLEAGASPGGSIYNNCSVPLTAARDGAVAIL 286
 DB 94 GAEVDLVKVGQTALYVAVVNGHLESTEILLLEAGADPNGRHRSRTPVYHAXRVRGRDIL 153

 QY 287 QELLDHGAENYKAKI-----PVWASNIAS---CSGPLYLAAYVGHLDVVRVRL 338
 DB 154 KALIRYGADVNVNHLTPDTPFPFSSRLTSLVVC--PLYISAAYHNLCQFRLLLQAGNP 211

 QY 339 DYNCTOGLLARVPR--PRTLLEICLHNCPEYIQLLIDFGANIYL---PSLSLDLTSQ 393
 DB 212 DFNCNGPVNTQEFYRGSPGCVMDAVLRHGCEAAFVSLVVEFGANLNVKWSLGPARGR 271

 QY 394 ---DDKGIALLOARATPRSLLSQVRLVVRRL 423
 DB 272 RKMDPEALQVFKEARSIPRTLLSLCRVAVRRAL 304

 RESULT 9
 US-09-908-805B-25
 ; Sequence 25, Application US/09908805B
 ; Patent No. US20020147307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HILTON, Douglas J
 ; APPLICANT: ALEXANDER, Warren S
 ; APPLICANT: VINEY, Elizabeth M
 ; APPLICANT: WILLSON, Tracey A
 ; APPLICANT: RICHARDSON, Rachael T
 ; APPLICANT: STARR, Robyn
 ; APPLICANT: NICHOLSON, Sandra E
 ; APPLICANT: METCALF, Donald
 ; APPLICANT: NICOLA, Nicos A
 ; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
 ; FILE REFERENCE: 10976Z
 ; CURRENT FILING DATE: 2001-07-19
 ; PRIOR FILING DATE: 1997-10-31
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 350
 ; TYPE: PR1
 ; ORGANISM: Mouse
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (167)
 ; OTHER INFORMATION: Xaa is unsure
 ;
 ; US-09-908-805B-25

 Query Match 20.5%; Score 480; DB 10; Length 350;
 Best Local Similarity 40.7%; Pred. No. 8.5e-41;
 Matches 111; Conservative 51; Mismatches 90; Indels 21; Gaps 7;

 QY 170 LNOAVYDNDSYTLQDLQERKFRFNSRSGW---GVPGTPLRLAASVGHLSCLQVLLAH 226
 DB 34 LHDAAYVGDQLTLRLNLQESYRSRINEKSVWCCGMLPCTPLRIAATAAGHCNCVDFLRK 93

 QY 227 GADVSLDVKAQOTPLFTAVSHGHLDVVRVLEAGASPGGSIYNNCSVPLTAARDGAVAIL 286
 DB 94 GAEVDLVKVGQTALYVAVVNGHLESTEILLLEAGADPNGRHRSRTPVYHAXRVRGRDIL 153

 QY 287 QELLDHGAENYKAKI-----PVWASNIAS---CSGPLYLAAYVGHLDVVRVRL 338
 DB 154 KALIRYGADVNVNHLTPDTPFPFSSRLTSLVVC--PLYISAAYHNLCQFRLLLQAGNP 211

 QY 339 DYNCTOGLLARVPR--PRTLLEICLHNCPEYIQLLIDFGANIYL---PSLSLDLTSQ 393
 DB 212 DFNCNGPVNTQEFYRGSPGCVMDAVLRHGCEAAFVSLVVEFGANLNVKWSLGPARGR 271

 QY 394 ---DDKGIALLOARATPRSLLSQVRLVVRRL 423
 DB 272 RKMDPEALQVFKEARSIPRTLLSLCRVAVRRAL 304

 RESULT 9
 US-09-908-805B-25
 ; Sequence 25, Application US/09908805B
 ; Patent No. US20020147307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HILTON, Douglas J
 ; APPLICANT: ALEXANDER, Warren S
 ; APPLICANT: VINEY, Elizabeth M
 ; APPLICANT: WILLSON, Tracey A
 ; APPLICANT: RICHARDSON, Rachael T
 ; APPLICANT: STARR, Robyn
 ; APPLICANT: NICHOLSON, Sandra E
 ; APPLICANT: METCALF, Donald
 ; APPLICANT: NICOLA, Nicos A
 ; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
 ; FILE REFERENCE: 10976Z
 ; CURRENT FILING DATE: 2001-07-19
 ; PRIOR FILING DATE: 1997-10-31
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 350
 ; TYPE: PR1
 ; ORGANISM: Mouse
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (167)
 ; OTHER INFORMATION: Xaa is unsure
 ;
 ; US-09-908-805B-25

 Query Match 20.5%; Score 480; DB 10; Length 350;
 Best Local Similarity 40.7%; Pred. No. 8.5e-41;
 Matches 111; Conservative 51; Mismatches 90; Indels 21; Gaps 7;

 QY 170 LNOAVYDNDSYTLQDLQERKFRFNSRSGW---GVPGTPLRLAASVGHLSCLQVLLAH 226
 DB 34 LHDAAYVGDQLTLRLNLQESYRSRINEKSVWCCGMLPCTPLRIAATAAGHCNCVDFLRK 93

 QY 227 GADVSLDVKAQOTPLFTAVSHGHLDVVRVLEAGASPGGSIYNNCSVPLTAARDGAVAIL 286
 DB 94 GAEVDLVKVGQTALYVAVVNGHLESTEILLLEAGADPNGRHRSRTPVYHAXRVRGRDIL 153

 QY 287 QELLDHGAENYKAKI-----PVWASNIAS---CSGPLYLAAYVGHLDVVRVRL 338
 DB 154 KALIRYGADVNVNHLTPDTPFPFSSRLTSLVVC--PLYISAAYHNLCQFRLLLQAGNP 211

 QY 339 DYNCTOGLLARVPR--PRTLLEICLHNCPEYIQLLIDFGANIYL---PSLSLDLTSQ 393
 DB 212 DFNCNGPVNTQEFYRGSPGCVMDAVLRHGCEAAFVSLVVEFGANLNVKWSLGPARGR 271

 QY 394 ---DDKGIALLOARATPRSLLSQVRLVVRRL 423
 DB 272 RKMDPEALQVFKEARSIPRTLLSLCRVAVRRAL 304

 RESULT 9
 US-09-908-805B-25
 ; Sequence 25, Application US/09908805B
 ; Patent No. US20020147307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HILTON, Douglas J
 ; APPLICANT: ALEXANDER, Warren S
 ; APPLICANT: VINEY, Elizabeth M
 ; APPLICANT: WILLSON, Tracey A
 ; APPLICANT: RICHARDSON, Rachael T
 ; APPLICANT: STARR, Robyn
 ; APPLICANT: NICHOLSON, Sandra E
 ; APPLICANT: METCALF, Donald
 ; APPLICANT: NICOLA, Nicos A
 ; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
 ; FILE REFERENCE: 10976Z
 ; CURRENT FILING DATE: 2001-07-19
 ; PRIOR FILING DATE: 1997-10-31
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 350
 ; TYPE: PR1
 ; ORGANISM: Mouse
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (167)
 ; OTHER INFORMATION: Xaa is unsure
 ;
 ; US-09-908-805B-25

 Query Match 20.5%; Score 480; DB 10; Length 350;
 Best Local Similarity 40.7%; Pred. No. 8.5e-41;
 Matches 111; Conservative 51; Mismatches 90; Indels 21; Gaps 7;

 QY 170 LNOAVYDNDSYTLQDLQERKFRFNSRSGW---GVPGTPLRLAASVGHLSCLQVLLAH 226
 DB 34 LHDAAYVGDQLTLRLNLQESYRSRINEKSVWCCGMLPCTPLRIAATAAGHCNCVDFLRK 93

 QY 227 GADVSLDVKAQOTPLFTAVSHGHLDVVRVLEAGASPGGSIYNNCSVPLTAARDGAVAIL 286
 DB 94 GAEVDLVKVGQTALYVAVVNGHLESTEILLLEAGADPNGRHRSRTPVYHAXRVRGRDIL 153

 QY 287 QELLDHGAENYKAKI-----PVWASNIAS---CSGPLYLAAYVGHLDVVRVRL 338
 DB 154 KALIRYGADVNVNHLTPDTPFPFSSRLTSLVVC--PLYISAAYHNLCQFRLLLQAGNP 211

 QY 339 DYNCTOGLLARVPR--PRTLLEICLHNCPEYIQLLIDFGANIYL---PSLSLDLTSQ 393
 DB 212 DFNCNGPVNTQEFYRGSPGCVMDAVLRHGCEAAFVSLVVEFGANLNVKWSLGPARGR 271

 QY 394 ---DDKGIALLOARATPRSLLSQVRLVVRRL 423
 DB 272 RKMDPEALQVFKEARSIPRTLLSLCRVAVRRAL 304

 RESULT 9
 US-09-908-805B-25
 ; Sequence 25, Application US/09908805B
 ; Patent No. US20020147307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HILTON, Douglas J
 ; APPLICANT: ALEXANDER, Warren S
 ; APPLICANT: VINEY, Elizabeth M
 ; APPLICANT: WILLSON, Tracey A
 ; APPLICANT: RICHARDSON, Rachael T
 ; APPLICANT: STARR, Robyn
 ; APPLICANT: NICHOLSON, Sandra E
 ; APPLICANT: METCALF, Donald
 ; APPLICANT: NICOLA, Nicos A
 ; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
 ; FILE REFERENCE: 10976Z
 ; CURRENT FILING DATE: 2001-07-19
 ; PRIOR FILING DATE: 1997-10-31
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 350
 ; TYPE: PR1
 ; ORGANISM: Mouse
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (167)
 ; OTHER INFORMATION: Xaa is unsure

QY 160 -----EDT-----DTEKQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHLEHIKILAGALRIADKVESGKTSVVVHCSDGWDRTAQLTSLAMLMDGY 365
QY 193 RFINS-----RSGWGVPGTPLRLAASGYHLSCLQVLLAHG-----ADVDSLVDKVAQTPLFT 243
Db 366 RTIRGEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
QY 244 AVSHGHLDVCV 253
Db 408 Q-----FIDCV 413

RESULT 14
US-10-370-659-11
; Sequence 11, Application US/10370659
; Publication No. US20030157681A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO01174DIV
; CURRENT APPLICATION NUMBER: US/10/370,659
; CURRENT FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-370-659-11

Query Match 12.5%; Score 293; DB 12; Length 458;
Best Local Similarity 27.7%; Pred. No. 3.5e-21;
Matches 86; Conservative 49; Mismatches 91; Indels 84; Gaps 13;

QY 3 ESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVWPKSVTLGTVVGSSKFR 62
Db 129 ENGKLYDPLLEVRROGIPNESWRTIKINERVELCDTYPALLVVPANIPDEELKRVASFR 188
QY 63 SKERPVLVLYKENNAACRCQSPLSGFY--TRCVDELLELAISQTNPGSQFMVYVDTR 121
Db 189 SGRIPVLSWIHPESQATITRCSQPMVGVSGKSKEDKYLQAIMDSNAQSHKIFIFDAR 248
QY 122 PKI-----WHFLVLMIRIVLQAKNLM--DITKIFSLLOPDKEE--- 159
Db 249 PSYNAVANKAGGGESEDAYQNAELVF---LDIHNHVMRESLRKLEIVYPNIEETHW 305
QY 160 -----EDT-----DTEKQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHLEHIKILAGALRIADKVESGKTSVVVHCSDGWDRTAQLTSLAMLMDGY 365
QY 193 RFINS-----RSGWGVPGTPLRLAASGYHLSCLQVLLAHG-----ADVDSLVDKVAQTPLFT 243
Db 366 RTIRGEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
QY 244 AVSHGHLDVCV 253
Db 408 Q-----FIDCV 413

RESULT 15
US-09-919-039-134
; Sequence 134, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113

; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 134
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 1823159CDI
US-09-919-039-134

Query Match 11.3%; Score 264; DB 11; Length 603;
Best Local Similarity 26.7%; Pred. No. 5.3e-18;
Matches 80; Conservative 46; Mismatches 106; Indels 68; Gaps 12;

QY 5 GWKLDIPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVWPKSVTLGTVVGSSKFRSK 64
Db 163 GWTYNPVEEYRROGLPNHHWRTITFINKCYELCDTYPALLVVPYRASDDDLRRVATFRSR 222
QY 65 ERVPVLVLYKENNAACRCQSPLSGFY--TRCVDELLELAISQTNPGSQFMVYVDTRPK 123
Db 223 NRIPVLSWIHPENKTVIVRCQPLVMSGSKRNKDEKYLDIRETNKQISKUITYDARPS 282
QY 124 I-----WHFLVLMIRIVLQAKNLM--DITKIFSLLOPDKEE--- 159
Db 283 VNAVANKATGGGYESDDAYHNAELFF---LDIHNHVMRESLRKLEIVYPNVEESHWLS 339
QY 160 ---EDTDEE--KQALNQAVYDNDYTLQDLRQERYKRPINRSRGVPGTPLRLA--- 211
Db 340 SLESTHLEHIKILVLTGATQVADK-----VSSGKSSVLVHCSDGWDRTAQLTSLAMLML 393
QY 212 -----ASYGHLSCLOVLLAHGADVSLDVKAQTPLFTAVSHGHLDVCV 253
Db 394 DSFYRSIEGFEILVQKEWISFGHKFASRI--GHG-DKNHTDAD--RSPILQ-----FIDCV 445

Search completed: December 1, 2003, 07:53:52
Job time : 540 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:39:07 ; Search time 4237 Seconds
(without alignments)
4335.242 Million cell updates/sec

Title: US-09-941-831A-20

Perfect score: 2343

Sequence: 1 MRESGWKLIDPISDFGRMG.....QAINQLDIPPMISYLKQL 449

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454913386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:

-MODE=frame+pn.model -DEV=xl
-O=/cgn2_1/USPTO_spool_p/US09941831/runat_01122003_073527_7829/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=200 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pflo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09941831@cgn_1_1_3745@runat_01122003_073527_7829 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1690	72.1	1222	6	AX418418 Sequence
2	1681	71.7	1269	9	AK096896 Homo sapi
3	1629	69.5	1069	6	AX418416 Sequence
C	1606.5	68.6	104534	9	AL356317 Human DNA
	1599	68.2	927	9	AF403030 Homo sapi
6	1459	62.3	1262	10	AF403040 Mus muscu
C	1391	59.4	191093	10	AL671765
	1376.5	58.7	194079	2	AC109692 Rattus no
9	1376.5	58.7	225982	2	AC108644
10	899.5	38.4	240816	5	AL844570 Zebrafish
11	675.5	28.8	2653	9	BC012399 Homo sapi
12	642.5	27.4	1345	6	AX406008 Sequence
13	499	21.3	1008	9	AF156777 Homo sapi
14	499	21.3	1340	9	BC014528 Homo sapi
15	481	20.5	1008	10	AF155352 Mus muscu
16	481	20.5	5892	10	AK122456 Mus muscu
17	480	20.5	2019	6	AR262497 Sequence
18	480	20.5	2019	6	AX022304 Sequence
19	480	20.5	2019	6	AX030771 Sequence
20	480	20.5	2019	6	BD007458 Remedies
21	454	19.4	6518	9	AB032972 Homo sapi
22	434.5	18.5	2354	5	BC053411 Danio rer
23	431.5	18.4	1312	10	BC049699 Mus muscu
24	424.5	18.1	2353	6	AX086038 Sequence
25	421	18.0	3065	10	BC020019 Mus muscu
26	409	17.5	2337	9	AK093237 Homo sapi
27	409	17.5	4069	9	BC040012 Homo sapi
28	408	17.4	1866	9	AF406619 Homo sapi
29	400	17.1	2754	5	BC044359 Danio rer
30	397.5	17.0	2817	10	BC032254 Mus muscu
31	395	16.9	4131	9	HSB803324
32	394	16.8	2482	6	AX399878 Sequence
33	378.5	16.2	3059	3	BT003304 Drosophil
34	378.5	16.2	3209	3	AF181629 Drosophil
35	335	14.3	77274	2	AC005672 Drosophil
36	335	14.3	154895	2	AC020329 Drosophil
37	335	14.3	181955	3	AC008304 Drosophil
38	335	14.3	218565	3	AC009257 Drosophil
C	335	14.3	299331	3	AE003460 Drosophil
	313.5	13.4	42724	3	CEP53A2
40	305.5	13.0	1214	3	AK113728
41	297	12.7	1966	6	BD156744 Primer fo
42	297	12.7	1966	9	AK001816 Homo sapi
43	296.5	12.7	2397	9	BC052990 Homo sapi
44	296.5	12.7	3158	9	BC040432 Homo sapi
45	296.5	12.7	3158	9	BC040432 Homo sapi

ALIGNMENTS

RESULT 1

```

AX418418
LOCUS AX418418 1222 bp DNA linear PAT 18-JUN-2002
DEFINITION Sequence 13 from Patent WO0206329.
ACCESSION AX418418
VERSION AX418418.1 GI:21523342
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE
1 Rastelli, L., Shimkets, R.A., Zerkhusen, B., Malyankar, U.M. and Padigaru, M.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0206329-A 13 24-JAN-2002;
Curagen Corporation (US)
FEATURES
source
1.1222
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="NOVX nucleic acid with homology to ankyrin repeat-containing- (ASB-1) like protein"
BASE COUNT 276 a 352 c 278 g 316 t
ORIGIN
Alignment Scores:
Pred. No.: 4.57e-147 Length: 1222
Score: 1690.00 Matches: 326
Percent Similarity: 99.70% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 72.13% Indels: 0
DB: 6 Gaps: 0
US-09-941-831A-20 (1-449) x AX418418 (1-1222)
QY 122 ProValIleThrHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMet 141
DB 102 CCACAGATCGCATTTCTTGGCTCATATGAGATAGTCTCCATATGACCAAGATG 161
QY 142 AsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAsp 161
DB 162 AACCTCATGGACATCACCAAGATCTTCTCCCTCGACGCCGACCAAGGAGGAGGAC 221
QY 162 ThrAspThrGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThr 181
DB 222 ACTGACACAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACACGACTCCTATAC 281
QY 182 LeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyr 201
DB 282 TTGGACCACTTTTGGCCAGGAGCGTTTCAACAGTTTTCATCACAGCAGGAGTGCTGG 341
QY 202 GlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGln 221
DB 342 GGTGTTCTCGGACACCTTGGCTTGGCTGCTTCTTATGGCCACTTGAGCTGTTGCAA 401
QY 222 ValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeu 241
DB 402 GTCCCTCTTAGCCCATCGTGTATGTGACAGCTTGGATGTCAAGGCACAGACGCCACTT 461
QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261
DB 462 TTCACCTGCTGTACGTATCGCCATCTGGACTGTGTACGTGTGCTTTGGAGCTGTGCTG 521
QY 262 SerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGly 281
DB 522 TCTCCTGGTGTAGCATCTACAACTGTCTCCGCTGCTCACAGCTGCCGTGATAGT 581
QY 282 AlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLys 301
DB 582 GCTGTGTCTATCTCGCAGAGCTCTTAGACCATGTGTGAGAGCCCAACGTCAAGCTAA 641
QY 302 LeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaLys 321

```

```

DB 642 CTACCAGTCTGGCATCAAAACATAGCTTCTATGTTCTGGCCCCCTCTATTGTCGGCAGTC 701
QY 322 TyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn 341
DB 702 TACGGGCACCTGGACTGTTTCCGCTGCTTTTGTCTCCACGGGGCAGACCTGACTACAC 761
QY 342 CysThrAspGlnGlyLeuLeuAlaArgValProAspProArgThrLeuLeuGluLeuCys 361
DB 762 TGCACTGACCGGGCTATTTGGCTGTGTCCCAAGACCCCGCACCTCTCTTGAATCTGC 821
QY 362 LeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381
DB 822 CTCATCATTAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGGTCTAATATC 881
QY 382 TyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeu 401
DB 882 TACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGGCATTCATTTGCTG 941
QY 402 LeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgArg 421
DB 942 CTACAGGCCGAGGACCTCCACGGTCACTTCTATCACAGTCCGTTAGTCTGCCAGA 1001
QY 422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeu 441
DB 1002 GCCTTGTGCCAGGCTGGCCAGCCACAGCATCAACAGCTGGATATTCCTCCCATGTTG 1061
QY 442 IleSerTyrLeuLysHisGlnLeu 449
DB 1062 ATTAGCTACCTAAAACACCAACTG 1085
RESULT 2
LOCUS AK096896 1269 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ39577 fis, clone SKMUS2002920, weakly similar to Homo sapiens ASB-1 protein mRNA.
ACCESSION AK096896
VERSION AK096896.1 GI:21756490
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K. and Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
AUTHORS Isogai, T. and Yamamoto, J.
JOURNAL Direct Submission
COMMENT Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES
source
1.1269
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SKMUS2002920"

```

CDS		<pre> /tissue_type="skeletal muscle" /clone_lib="SKM052" /note="cloning vector: pME18SFL3" 174..1130 /note="unnamed protein product" /codon_start=1 /protein_id="BAC04888.1" /db_xref="GI:21756491" /translat="MRIVLQAKNLMIDITKIFSLQPKDEEDTDEERKALNQAVSL DNDSTYDQLLRERYKRFINSRSGMGVPTPLRLAAASYGHLSCLOVLLAHGADVLSL DVAQATPLFTAVSHGHUDCVRLLEAGASPGSGTYNNCSPLVLTAAQDGAVALIQEPLD HGAEANVKALPVWASNIASCSPLXLAAYVGHLDLCPRLILLHAARDGAYNCTDQGLLA RVPRPTLLIEICLHNCPEPYIQLLIDFGANIVLPSSLTSDODDKGIALLLQARRAT PRSLSSQRLVVRALCAQQAQAINOLDIPPLMLISYLKQL" </pre>									
BASE COUNT	288 a	370 c	305 g	306 t							
ORIGIN											
Alignment Scores:											
Pred. No.:	3..29e-146	Length:	1269								
Score:	1681.00	Matches:	325								
Percent Similarity:	99.69%	Conservative:	0								
Best Local Similarity:	99.69%	Mismatches:	1								
Query Match:	71.75%	Indels:	0								
DB:	9	Gaps:	0								
US-09-941-831A-20 (1-449) x AK096896 (1-1269)											
Qy	124	IleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMetAsnLeu	143								
Db	150	ATCTGGCATTTCTTGTGCTCATPAATGAGAAATAGTTCTCCAAATTAGCAAGATGACCTC	209								
Qy	144	MetAspIleThrIlyPheSerLeuLeuGlnProAspIysGluGluAspThrAsp	163								
Db	210	ATGACATACCAAGATCTTCTCCTCTCTGCAGCCGACAGGAGGAGGACACTGAC	269								
Qy	164	ThrGluGluIysGlnAlaLeuAsnGlnAlaValIlyrAspAsnAspSerTyrThrLeuAsp	183								
Db	270	ACAGAGGAGAGCAGGCTCTCAATCAACAGCAGTGTATGACACGACTCCTATACTTTGGAC	329								
Qy	184	GlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyrGlyVal	203								
Db	330	CAGCTTTGGCCAGGACGGTTACAAACGTTTCAACACAGCAGAGTGGCTGGGGGTGT	389								
Qy	204	ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu	223								
Db	390	CCTGGGACACCTTTGGCTTGGCTGCTTCTATGGCCACTTGAGCTGTTTGCAGTCTCTC	449								
Qy	224	LeuAlaHisGlyAlaAspValAspSerLeuAspValIlyAlaGlnThrProLeuPheThr	243								
Db	450	TTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAGGCGACAGCGCCACTTTTCACT	509								
Qy	244	AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro	263								
Db	510	GCTGTAGTCATGGCCATCTGGACTGTGAGTGTGCTTTTGGAAAGCTGGTCTCTCTCT	569								
Qy	264	GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal	283								
Db	570	GGTGGTAGCATCTACAAACACTGTTCTCCCGTGCTCACAGCTGCCCGTGATGGTGTGT	629								
Qy	284	AlaIleLeuGlnGluLeuAspHisGlyAlaGluAlaAsnValIlyAlaLysLeuPro	303								
Db	630	GCTATCTCGAGGAGCCCTAGACCATGGTGCAGAGGCCAACGCTCAAAAGCTAAACTACCA	689								
Qy	304	ValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGly	323								
Db	690	GTCCTGGGCATCAAAACATAGCTTTCATGTTCTGGCCCCCTCTATTTGGCGCGAGTCTACGG	749								
Qy	324	HisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThr	343								
Db	750	CACCTGACTGTTTCCGCCTGCTTTTGCTCCACGGGCGAGNCCCTGACTCAACTGCACT	809								
Qy	344	AspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHis	363								

Db	810	GACCAGGGCCCTATTGGCTGTGTGCCAAGACCCCGACCCCTCTTGAATCTCGCTCCAT	869
Qy	364	HiAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeu	383
Db	870	CATAATGTGAGCCAGATATATCCAGCTGTTAATCGATTTGGTGTCTAATATCTACCTT	929
Qy	384	ProSerLeuSerLeuAspLeuThrSerGlnAspAspIysGlyIleAlaLeuLeuGln	403
Db	930	CCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGCATTTGCAATGCTGTACAG	989
Qy	404	AlaArgAlaThrProArgSerLeuSerGlnValArgLeuValValArgAtgAlaLeu	423
Db	990	GCCCGAGCACTCCAGGTCACTTATACAGGTCCGTTAGTCTCCGAGACCTTG	1049
Qy	424	CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeuIleSer	443
Db	1050	TGCCAGGCTGGCCAGCCCAAGCCATCAACAGCTGGATATCTCTCCCATGTTGATTAGC	1109
Qy	444	TyrLeuIysHisGlnLeu	449
Db	1110	TACCTAAACACCAACTG	1127
RESULT 3			
AX418416	AX418416	1069 bp	DNA linear PAT 18-JUN-2002
LOCUS	Sequence 11 from Patent WO0206329.		
DEFINITION	AX418416		
ACCESSION	AX418416.1	GI:21523341	
VERSION			
KEYWORDS	unidentified		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1.		
AUTHORS	Rastelli, L., Shimkets, R.A., Zerhusen, B., Malyankar, U.M. and Padigaru, M.		
TITLE	Human polynucleotides and polypeptides encoded thereby		
JOURNAL	Patent: WO 0206329-A 11 24-JAN-2002;		
FEATURES	Curagen Corporation (US)		
source	Location/Qualifiers		
	1..1069		
	/organism="unidentified"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32644"		
	/notes="NOVX nucleic acid with homology to ankyrin repeat-containing- (ASP-1) like protein"		
BASE COUNT	237 a 258 c 257 g 277 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.72e-141	Length:	1069
Score:	1629.00	Matches:	318
Percent Similarity:	97.55%	Conservative:	0
Best Local Similarity:	97.55%	Mismatches:	0
Query Match:	69.53%	Indels:	8
DB:	6	Gaps:	1
US-09-941-831A-20 (1-449) x AX418416 (1-1069)			
Qy	132	MetArgIleValLeuGlnLeuAlaLysMetAsnLeuMetAspIleThrLysIlePheSer	151
Db	81	ATGAGAATAAGTTCTCCAAATTAGCAAGATGAACCTCATGGACATCAACAAGATCTCTCC	140
Qy	152	LeuLeuGlnProAspIysGluGluGluAspThrAspThrGluGluLysGlnAlaLeuAsn	171
Db	141	CTCCTGAGCCGCCACAGGAGGAGGAGCACCTGACACAGGAGACAGGCTCTCAAT	200
Qy	172	GlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyr	191
Db	201	CAAGCAGTGTATGACAAACGACTCTATATCTTGACCAGCTTTTGGCCAGGAGCGTTAC	260
Qy	192	LysArgPheIleAsnSerArgSerGlyTyrIleValProGlyThrProLeuArgLeuAla	211
Db	261	AAACGTTTCATCAACAGCAGAGTGGCTGGGGTTCCTGGACACCTTGGCTGGCTGGCT	320

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known, annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP11-284B18 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pSPACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-284B18. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-284B18 is at 1 in this sequence. The true left end of clone RP4-710J4 is at 104435 in this sequence. The true right end of clone RP11-403E24 is at 63843 in this sequence.

FEATURES source

```

1. .104534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="q12-13.1"
/clonene="rp11-284B18"
/clonelib="RPCL11.1"
1. .4482
/notes="HERVL repeat: matche
repeats=489.5037
/notes="MLT2B repeat: matc
repeats=5038.5133
/notes="L1MA7 repeat: matc
repeats=5329.5398
/notes="L2 repeat: matche
repeats=6136.6274
/notes="M1R repeat: matche
repeats=6859.7220
/notes="MLT1A1 repeat: matc
repeats=8549.8744
/notes="LTR33 repeat: matc
repeats=12340.12716
/notes="M1R repeat: matche
repeats=12732.13094
/notes="L2 repeat: matche
repeats=13311.13725
/notes="L2 repeat: matche
repeats=13364.14156
/notes="M1R repeat: matche
repeats=14257.14330
/notes="M1R repeat: matche
repeats=14386.14524
/notes="M1R repeat: matche
repeats=15029.15467
/notes="Charlie4a repeat:
repeats=16483.17040
/notes="L1P4a repeat: matc
repeats=17775.18439
/notes="L1MD2 repeat: matc
repeats=18432.18880
/notes="L1M4 repeat: matc

```

QY	212	AlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp	231
Db	321	GCTTCTTATGGCCACTTTCAGCTGTTTTCAGAGTCTCTTAGCCCATGCTGCTGATGTTGAC	380
QY	232	SerLeuAspValIysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp	251
Db	381	AGCTTGAGATGTCAAGGCACAGACGCCACTTTTCACTGCTGTCAGTCATGAGCCATCTCGGAC	440
QY	252	CysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys	271
Db	441	TGTGTACGTGTCTTTTGGAAAGCTGGTGCCTCTCTCTGCTGGTAGCATCTACAACTGCT	500
QY	272	SerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAsp	291
Db	501	TCTCCCGTGTCTCAGCTGCGCGTGAATGCTGCTTGTCTATCTCTGTCAGGAGCTCCTAGAC	560
QY	292	HisGlyAlaGluAlaAsnValIysAlaIysLeuProValTrpAlaSerAsnIleAlaSer	311
Db	561	CATGGTGCAGAGGCCAACGTCAAAGCTTAAACTACAGTCTGGGATCAAAACATAGCTTCA	620
QY	312	CysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeu	331
Db	621	TGTTCTTGCCCCCTCTATTATTTGGCGCAGTCTACGGGCACCTGGACTGTTTCCGCGCTGCTT	680
QY	332	LeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgVal	351
Db	681	TTTGCTCACCGGGCAGACCCCTGACTACAACCTGCATCAGCAGGCGCTATTGGCTGCTGTC	740
QY	352	ProArgProArgThrLeuLeuGluIleCysLeuHisAsnCysGluProGluTyrIle	371
Db	741	CCAAGACCCCGCACCTCTTGAATCTGCTCCATCATTAATGTGAGCCAGAGTATATC	800
QY	372	GlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeuAspLeuThr	391
Db	801	CAGCTGTTAATCGATTTTGGTGTAAATATCTACCTTCCATCTCTCTCCCTTGCCTGACC	860
QY	392	SerGlnAspAspIysGlyIleAlaLeuLeuLeuGlnAlaArg	405
Db	861	TCACAAGATGATAAAGGCATTTGCATTCTCTCATAGCCCGAGGTGAGCTGTTCTTCTTCTT	920
QY	406	-----AlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeu	423
Db	921	GCTGTAGCCACTCCACGGTCACTTCTATCATCAGGTCCGTTTGTCTGTCGCGAGACCTTG	980
QY	424	CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeuIleSer	443
Db	981	TGCCAGGCTGGCCAGCCACAAGCCATCAACAGCTGGATATCTCTCCCATGTTGATTAGC	1040
QY	444	TyrLeuIysHisGlnLeu	449
Db	1041	TACCTAAACACCAACTG	1058
RESULT 4	AL356317/C		
LOCUS	AL356317	104534 bp	DNA linear PRI 02-FEB-2001
DEFINITION	Human DNA sequence from clone RP11-284B18 on chromosome Xq12-13.1, complete sequence.		
ACCESSION	AL356317		
VERSION	AL356317.8	GI:12666252	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Lawlor, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
COMMENT	requests: Clonerquest@sanger.ac.uk On Feb 5, 2001 this sequence version replaced gi:12578192. During sequence assembly data is compared from overlapping clones.		


```
Db 10836 TTCACTGCTGTAGTCAATCGCCATCTGAGCTGTGTGTGCTTTTGGAAAGCTGTGGCC 10777
Qy 262 SerProGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGly 281
Db 10776 TCTCCTGGTGTAGCATCTACAACAATGTTCTCCGCTGTCTACAGCTGCCGTGTAGGT 10717
Qy 282 AlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnVallyAlaLys 301
Db 10716 GCTGTTGCTATCTCGCAGGAGCTCTAGACCATGTTGCAGAGCCAAACGTCAAAGCTAAA 10657
Qy 302 LeuProValTrpAlaSerIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaVal 321
Db 10656 CTACCAAGTCTGGGCATCAACATAGCTTCATGTTCTGGCCCCCTCTATTGGCCGAGTC 10597
Qy 322 TyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn 341
Db 10596 TAGGGGACCTGGACCTGTTTCCGCCCTGCTTTTCTCCAGGGGACACCTGTACTACAAC 10537
Qy 342 CysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCys 361
Db 10536 TGCACCTGACCAAGGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTCTTGAATCTGC 10477
Qy 362 LeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIle 381
Db 10476 CTCATCATATAATTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGTGTCTAATATC 10417
Qy 382 TyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeu 401
Db 10416 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATTAAGCATTTGCTGCTG 10357
Qy 402 LeuGlnAlaArgAla ----- 406
Db 10356 CTACAGGCCCGAGG-TGAGTGTGTAACAGCCAGCCCTCATAGCCACCTGCTTGGAGGT 10298
Qy 406 ----- 406
Db 10297 CCCAGGGCTCTAAATCAGCTATAAGCAATCTCTGAGGTGCACTGGAGAAAACCTCCATC 10238
Qy 406 ----- 406
Db 10237 TTCCTCCCTCTCTGGGGCCCCAGTAGAGAAGGGGTTATGATTCATGTTAGAGAAA 10178
Qy 406 ----- 406
Db 10177 GGTCACTAAGATAAGTGTGGTGGGGTGGGAGCATTAGCCCAATTTCTGTTCTGGCT 10118
Qy 406 ----- 406
Db 10117 CTGCCAATGACTGTTTATTCTTTGGAAGGATAGAGCAATTAGGGCTGTAGTCCAGTGG 10058
Qy 406 ----- 406
Db 10057 AGAGTGGATGTAGGCTTTGTGGGAGAAATTTTCTGATTAATGCTCTCATGACCTGTTT 9998
Qy 407 -----ThrProArgSerLeuLeuSerGlnValArgLeuValValArgArg 421
Db 9997 CTTCTTGTCTAGCCACTCCAGCTCTCTATCATCAGCTCGTTTGTGCTCGCCGAGA 9938
Qy 422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeu 441
Db 9937 GCCTTGTGCGAGCTGGCCAGCCACAGCCATCAACAGCTGGATATTCTCTCCCATGTTG 9878
Qy 442 IleSerTyrLeuIleHisGlnLeu 449
Db 9877 ATTAGCTACTAAACACCAACTG 9854
RESULT 5
AF403030 927 bp mRNA linear PRI 02-JAN-2002
LOCUS
DEFINITION Homo sapiens ankyrin repeat domain-containing SOCS box protein
Asb-12 mRNA, complete cds.
ACCESSION AF403030
VERSION AF403030.1 GI:18034079
```

KEYWORDS
SOURCE

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 927)

Kile, B.T. and Nicola, N.A.

SOCS box proteins

Unpublished

2. (bases 1 to 927)

Kile, B.T. and Nicola, N.A.

Direct Submission

Submitted (27-JUN-2001) Cancer & Hematology, The Walter & Eliza

Hall Institute of Medical Research, Royal Melbourne Hospital,

Parkville, Victoria 3050, Australia

Location/Qualifiers

1. 927

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

1. 927

/codon_start=1

/product="ankyrin repeat domain-containing SOCS box

protein Asb-12"

/protein_id="AA157349.1"

/db_xref="GI:18034080"

/translation="NNLMDITKIFSLLOPDKEEDTDTEKQALNOAVYDNDSTLDDQ

LLRORYKRFINSRSGWVPGTPLRLAASYGHLSCLOVLLAHGADVSLDKVKAOTPLF

TAVSGHLDLCVRLVLEAGASPGSGSYNNCSFVLTAAAGDAVAILQELLDHGAENYKA

KLFWASASNYSCSGPLYLAAVYGHLDLFRLLLLHGDADPDYNTDGLLARVPRPRTLL

ETCLHNGEPEYIQLIDFGANILYPLSLDLTSDQDKGIALLQARATPRSLLSQVR

LVRRALCAQAGPOAINOLDIPPMILSYLKHQ"

BASE COUNT 203 a 274 c 216 g 234 t

ORIGIN

Alignment Scores:

Pred. No.: 8,56e-139 Length: 927

Score: 1599.00 Matches: 308

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 68.25% Indels: 0

DB: 9 Gaps: 0

US-09-941-831A-20 (1-449) x AF403030 (1-927)

Qy 141 MetAsnLeuMetAspIleThrIlySlePheSerLeuLeuGlnProAspLysGluGlu 160

Db 1 ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCTGAGCCGCCACAGGAGGAG 60

Qy 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180

Db 61 GACACTGACACAGAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGCTCTAT 120

Qy 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200

Db 121 ACTTTGGACCACTTTTGGCCAGGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGGC 180

Qy 201 TrpGlyValProGlyThrProLeuArgLeuAlaAspTyrGlyHisLeuSerCysLeu 220

Db 181 TGGGGTGTTCCTGGGACACCTTGGCTTGGCTTCTTATGGCCACTTGAGCTCTTTG 240

Qy 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240

Db 241 CAAGTCTCTTAGCCCATGCTGTATGTTGACAGCTTGGATGTCAAGGCACAGAGCCCA 300

Qy 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260

Db 301 CTTTTCACCTGCTCAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 360

Qy 261 AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280

Db 361 GCCTCTCTCTGGTGTAGCATCTACAACTGTTCTCCCGTGTCTACAGCTGCCCTGTAT 420

```

QY 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVallySala 300
Db 421 GGTGCTGTTGCTATCTCGCAGGAGCTCTTAGACCATGGTGCAGAGCCCAAGCT 480
QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320
Db 481 AAATACCAAGTCTGGGCATCAAAATAGCTTCATGTTCTGGCCCCCTCTATTGGCCGCA 540
QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
Db 541 GTCTAGGGCAGCTGAGTCTTCCGCTGCTTTGCTCCAGGGGAGACCCCTGACTAC 600
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
Db 601 AACTGCACTGACCCAGGCGCTATTGGCTGGTGTCCCAAGACCCCGCACCCTCTTGAATC 660
QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
Db 661 TGCCTCCATCAATATTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTGGTGCTAAT 720
QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
Db 721 ATCTACTTCCATCTCTCTCCCTTGACCTCAGCTCAAGATGATAAGGCATTTGCATTG 780
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
Db 781 CTGCTACAGGCCCGAGCCACTCACGGTCACTTCTATCACAGCTCGTTAGTCTCGCG 840
QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMet 440
Db 841 AGAGCCTTGTCCAGGCTGCCAGCCACAGCCATCAACAGCTGGATATCTCCCATG 900
QY 441 LeuIleSerTyrLeuLysHisGln 448
Db 901 TTGATTAGCTACCTAAACACCAA 924

RESULT 6
AF403040 1262 bp mRNA linear ROD 02-JAN-2002
LOCUS Mus musculus ankyrin repeat domain-containing SOCS box protein
DEFINITION
AF403040
ACCESSION
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE
AUTHORS Karyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Kile, B.T. and Nicola, N.A.
JOURNAL SOCS box proteins
REFERENCE
AUTHORS Kile, B.T. and Nicola, N.A.
JOURNAL Unpublished
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza
Hall Institute of Medical Research, Royal Melbourne Hospital,
Parkville, Victoria 3050, Australia
FEATURES
source
1. 1262
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
198..1124
/codon_start=1
/product="ankyrin repeat domain-containing SOCS box
protein Ab-12"
/protein_id="AA157359.1"
/db_xref="GI:18034100"
/translators="NLM:MDIAKIFSLQPEKEEDTGTGKQALNOAVYDNDPSCTLDH
LHQRKYPFINSRSGWIPGTPLRLAASGYHLNCVKVLEHGADVDSLDVRAQPLF
TAVSHGLECVRMLLBAGCPGSIYNNCSPLVTSRDLGAFALQELHGHAEANVKA
KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLYLGADPDYNCCTDQGLLRVPQPRLL

```

```

BASE COUNT 318 a 338 c 272 g 334 t
ORIGIN
Alignment Scores:
Pred. No.: 124e-125 Length: 1262
Score: 1459.00 Matches: 282
Percent Similarity: 92.94% Conservative: 21
Best Local Similarity: 85.50% Mismatches: 22
Query Match: 62.27% Indels: 2
DB: 10 Gaps: 1
US-09-941-831A-20 (1-449) x AF403040 (1-1262)
QY 124 IleTrpHisPheLeuValLeuIleMetArgIleValLeu--GlnLeuAlaLysMetAsn 142
Db 145 ATCTGGAATTTCTCTCTCTGCAATGAGATA-GTGTCTATTTCGAATAAACACCAAGATGAAC 203
QY 143 LeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAspThr 162
Db 204 CTCATGGATATCGCCAGATCTTCTCTCTCTGCAACCTGAAAGAGGAGGAGGACACT 263
QY 163 AspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeu 182
Db 264 GGCACCGGGGAAAAACAGGCTCTCAATCAAGCTGTATATGACATGACTCTCTGTACCCCTG 323
QY 183 AspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrpGly 202
Db 324 GACCACTTCTACACCGAGGAGCTTATAAACGGTTTCATCAACAGCAGGAGTGGGGT 383
QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
Db 384 ATACCTGGAACACCCCTTGGCTTGGCAGCTTCTTATGTCCTTAAATTTGTGAGGCTC 443
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
Db 444 CTCCTGGAACATGGTGTGATGTGATAGTTGGATGTCAAAGCACAACACCCACTTTTC 503
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
Db 504 ACTGCTGTGAGCCACGGTCTCATCTGGAGTGGTGAATGCTTTTAGAAGCTGGTGGCTGT 563
QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
Db 564 CTTAGTGGTAGCATCTACAAACAATGCTCTCTCTCTCCTCCTCCTCCTCCTCCTGATGGGCT 623
QY 283 ValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
Db 624 TTTGGCACTTTACAGGAGCTCTTAGGGCATGGTCCGAGGCTTAATGTCAAAGCTAAACTA 683
QY 303 ProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaValTyr 322
Db 684 CCAGTCTGGGCGTCAAAATATAGCTTCTATGCTGGCCCTCTATCTGCTGAGTCTAT 743
QY 323 GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCys 342
Db 744 GGGCACTTGTATGTTTCCGCTGCTTTTGTCTTATGGGGGAGATCCTCATTAACAATGC 803
QY 343 ThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeu 362
Db 804 ACTGACCGAGGCTTTTAAGTGTGTTCACAGCTCCGACACTCTCTTGAATCTGCCCTT 863
QY 363 HisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyr 382
Db 864 CATCAATATTGTGAGCAGAGTACATCCAGCTTTTAAATAGATTTTGGAGCTAAACATCTAC 923
QY 383 LeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeu 402
Db 924 CTTCCATCTCTCTCTGTGGACCAACTTCCAAAGATGATAAAGGCAATCAAAATTCGCTGA 983
QY 403 GlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAla 422

```

Db 984 CAAGCCCGAGCCTCCACGGTCTACTCTGTGTCCAGACCCGTTAGTTATTCGCCGAGATCC 1043
 Qy 423 LeuCysGlnAlaglyGlnProGlnAlaileAAsnGlnLeuAspIleProPrometLeuile 442
 Db 1044 CTCTGCCGGGCGAACCAAGTCACACAGCCACCGACAGCTGGATATCCCCCTGTGTGATT 1103
 Qy 443 SerTyrLeuIysHisGln 448
 Db 1104 AGCTACCTCAACATCAA 1121

 RESULT 7
 AL671765/c
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-100L23 on chromosome X, complete sequence.
 ACCESSION AL671765
 VERSION AL671765.6 GI:21666377
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 191093)
 Lloyd,D.
 Direct Submission
 Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 2, 2002 this sequence version replaced gi:21614613.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are compared with variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-100L23 is from the RP23-23 Mouse PAC Library constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6.

 FEATURES
 source Location/Qualifiers
 1. 191093
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-100L23"
 /clone_lib="RPCi-23"
 BASE COUNT 58421 a 38555 c 38626 g 55491 t
 ORIGIN

 Alignment Scores:
 Pred. No.: 2,34e-116 Length: 191093
 Score: 1391.00 Matches: 283
 Percent Similarity: 66.30% Conservative: 22
 Best Local Similarity: 61.52% Mismatches: 22

```
QY 405 ----- 405
Db 90162 ACTAAGATGAGGTATGTAGAAAGCATTTATACATTCTCTGGTCTTAGATCTCTTAATGAC 90103
QY 405 ----- 405
Db 90102 GGACCCCTTTGATGCCCTGATAGGCTCAGTATATGTTATCATGCAATTGGGAAGATGGTT 90043
QY 405 ----- 405
Db 90042 GAGACAGTCTTAGAAGGAGGAGGCGCATTTACTATATAGCCATGCGAGNATAGATGTACA 89983
QY 406 -----AlaThrPro 408
Db 89982 CTTGGAATTACACTGAGAATTGATGCTCATATCTGTTTCTCATTTGTTGCAGCCCACTCCA 89923
QY 409 ArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuLeuCysGlnIleGlyGln 428
Db 89922 CGGTCACTCTCTGCCAGACCCGTTTAGTTATCCGAGATCCCTCTGCCGGCGGAACACG 89863
QY 429 ProGlnAlaIleAsnGlnLeuAspIleProMetLeuIleSerTyrLeuIleValHisGln 448
Db 89862 TCACAGGCCACCGACCGAGCTGGATATCCCTGTTGTTAGTTAGTACCTCAACATCAA 89803

RESULT 8
AC109692 194079 bp DNA linear HTG 11-OCT-2002
LOCUS Rattus norvegicus clone CH230-268J15, *** SEQUENCING IN PROGRESS
DEFINITION ***
AC109692
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus.

REFERENCE
1 (bases 1 to 194079)
Muzny, D., Marie, Metzker, M., Lee, A., Brannon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gregg, G., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Pastermak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quintero, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
```

```
TITLE
JOURNAL
AUTHORS
REFERENCE
JOURNAL
AUTHORS
TITLE
COMMENT

Unpublished
2 (bases 1 to 194079)
Worley, K. C.
Direct Submission
Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 194079)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21737980.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQBH
Center clone name: CH230-268J15
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 167200 bases at least Q40
Consensus quality: 170449 bases at least Q30
Consensus quality: 172661 bases at least Q20
Estimated insert size: 186192; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194079: contig of 194079 bp in length.
* Location/Qualifiers
* 1..194079
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
```

```

/db_xref="taxon:10116"
/clone="CH230-268J15"
1..1257
/note="wgs_end_extension
clone_end:T7"
misc_feature
6621..7521
/note="clone_boundary
clone_end:T7"
site:Mbol
end sequence:RXAE056TJ"
110101..112444
/note="wgs_contig"
192195..193062
/note="clone_boundary
clone_end:Sp6
site:Mbol
end sequence:RXAE056TV"
BASE COUNT 48623 a 34416 c 35611 g 55333 t 20096 others
ORIGIN

Alignment Scores:
Pred. No.: 5.28e-115 Length: 194079
Score: 1376.50 Matches: 283
Percent Similarity: 62.76% Conservative: 22
Best Local Similarity: 58.23% Mismatches: 25
Query Match: 58.75% Indels: 156
DB: 2 Gaps: 3

US-09-941-831A-20 (1-449) x AC109692 (1-194079)
QY 119 AspThrArgProLysIleTyrHis----- 126
Db 152178 GAGAGTAAACATAGACTCTGGCAATGTAATAGTCTTAGCTTACACCTCCCTTTT 152237
QY 127 -----PheLeuValLeuIleMetArgIleValLeu---GlnLeu 138
Db 152238 CATTACTCTCAGATCGGGAATTTCTCTCTTTACAAATGAGATAGCTCTATTCCAATAA 152297
QY 139 AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
Db 152298 ACCAGATGAACTCATGATATCGCAAGATCTTCTCTTCTGCAACGAGAAAGAG 152357
QY 159 GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
Db 152358 GAGGAGGAGCGCCGACACAGCGGAAAGAGGCTCTCAATCAAGCTGTATATGACATGAC 152417
QY 179 SerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
Db 152418 TCTTGTAACCTTGACACCACTTCTACACCGGAAAGCTTATAAAGCTTATCAACAGCAGG 152477
QY 199 SerGlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
Db 152478 AGTGGCTGGGTATACCTGGAACACCTTTCGGCTTGGCAGCTCTTATGGTCACTTAGAT 152537
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
Db 152538 TGTGTCAAGAGTCTCTCTGGAACATGCTGCTGATGTTGATAGCTTGGATGTCAAGCACAA 152597
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
Db 152598 ACACCACTTTTCACTGCTGTAGTACGCTATCTGAGGATGTTGAGAGTACTTTTAGAA 152657
QY 259 AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
Db 152658 GCTGGTGGCTGCTCTAGTGTAGCATCTACAAATGTCTCTCTCTCTCTCTCTCTCTCTCT 152717
QY 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
Db 152718 CGTGTGGGGCCCTTCCCATCTTACAGGAGCTCTTAGGGCATGGTGCAGAGCTAATGTC 152777
QY 299 LysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeu 318
Db 152778 AAGGCGAAATCTGCAGTCTGGGCATCAAAACATAGCTTATGTTCTGGCCCCCTCTATTG 152837

/db_xref="taxon:10116"
/clone="CH230-268J15"
1..1257
/note="wgs_end_extension
clone_end:T7"
misc_feature
6621..7521
/note="clone_boundary
clone_end:T7"
site:Mbol
end sequence:RXAE056TJ"
110101..112444
/note="wgs_contig"
192195..193062
/note="clone_boundary
clone_end:Sp6
site:Mbol
end sequence:RXAE056TV"
BASE COUNT 48623 a 34416 c 35611 g 55333 t 20096 others
ORIGIN

Alignment Scores:
Pred. No.: 5.28e-115 Length: 194079
Score: 1376.50 Matches: 283
Percent Similarity: 62.76% Conservative: 22
Best Local Similarity: 58.23% Mismatches: 25
Query Match: 58.75% Indels: 156
DB: 2 Gaps: 3

US-09-941-831A-20 (1-449) x AC109692 (1-194079)
QY 119 AspThrArgProLysIleTyrHis----- 126
Db 152178 GAGAGTAAACATAGACTCTGGCAATGTAATAGTCTTAGCTTACACCTCCCTTTT 152237
QY 127 -----PheLeuValLeuIleMetArgIleValLeu---GlnLeu 138
Db 152238 CATTACTCTCAGATCGGGAATTTCTCTCTTTACAAATGAGATAGCTCTATTCCAATAA 152297
QY 139 AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
Db 152298 ACCAGATGAACTCATGATATCGCAAGATCTTCTCTTCTGCAACGAGAAAGAG 152357
QY 159 GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
Db 152358 GAGGAGGAGCGCCGACACAGCGGAAAGAGGCTCTCAATCAAGCTGTATATGACATGAC 152417
QY 179 SerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
Db 152418 TCTTGTAACCTTGACACCACTTCTACACCGGAAAGCTTATAAAGCTTATCAACAGCAGG 152477
QY 199 SerGlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
Db 152478 AGTGGCTGGGTATACCTGGAACACCTTTCGGCTTGGCAGCTCTTATGGTCACTTAGAT 152537
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
Db 152538 TGTGTCAAGAGTCTCTCTGGAACATGCTGCTGATGTTGATAGCTTGGATGTCAAGCACAA 152597
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
Db 152598 ACACCACTTTTCACTGCTGTAGTACGCTATCTGAGGATGTTGAGAGTACTTTTAGAA 152657
QY 259 AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
Db 152658 GCTGGTGGCTGCTCTAGTGTAGCATCTACAAATGTCTCTCTCTCTCTCTCTCTCTCTCT 152717
QY 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
Db 152718 CGTGTGGGGCCCTTCCCATCTTACAGGAGCTCTTAGGGCATGGTGCAGAGCTAATGTC 152777
QY 299 LysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeu 318
Db 152778 AAGGCGAAATCTGCAGTCTGGGCATCAAAACATAGCTTATGTTCTGGCCCCCTCTATTG 152837

AC108644
Rattus norvegicus clone CH230-59AG, *** SEQUENCING IN PROGRESS ***,
8 unordered pieces.
AC108644
AC108644.5 GI:24818595
HTG, HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 225982)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

```


Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Garregeorgis, E., Geard, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivett, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorusnew, L., Louleeged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morris, K., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okunolu, G., Olarnpungagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 225982)
Worley, K.C.

Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 225982)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23267762.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GPSJ
Center clone name: CH230-59A6

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 211822 bases at least Q40
Consensus quality: 215372 bases at least Q30
Consensus quality: 217787 bases at least Q20
Estimated insert size: 216316; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 32687: contig of 32687 bp in length
* 32688 32787: gap of unknown length
* 32788 102606: contig of 69819 bp in length
* 102607 102706: gap of unknown length
* 102707 123303: contig of 20597 bp in length
* 123304 123403: gap of unknown length
* 123404 123409: contig of 88306 bp in length
* 123410 212409: gap of unknown length
* 212410 213673: contig of 1264 bp in length
* 213674 213773: gap of unknown length
* 213774 215049: contig of 1176 bp in length
* 215050 217109: contig of 2060 bp in length
* 217110 217209: gap of unknown length
* 217210 225982: contig of 8773 bp in length.

FEATURES

source

1..225982
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-59A6"
complement(4374..5073)
/note="clone_boundary
clone_end:T7
site:ECORI
end sequence:BH331578"
32788..35656
/note="wgs_contig"
81431..82225
/note="clone_boundary
clone_end:Sp6
site:ECORI
end sequence:BH331579"

misc_feature

BASE COUNT 65353 a 44744 c 44903 g 64320 t 6662 others
ORIGIN

misc_feature

Alignment Scores:
Pred. No.: 6.49e-115 Length: 225982
Score: 1376.50 Matches: 283
Percent Similarity: 62.76% Conservative: 22
Best Local Similarity: 58.23% Mismatches: 25
Query Match: 58.75% Indels: 156
DB: 2 Gaps: 3

misc_feature

US-09-941-831A-20 (1-449) x AC108644 (1-225982)


```
QY 119 AspThrArgProLysLeuTyrHis-----126
Db 129983 GAGAGTAAACATAGAGTCTGGCATTTGATAATGATGCTTAGTCTTACACCCCTCCCTTTTT 130042
QY 127 -----PheLeuValLeuLeuMetArgIleValLeu---GlnLeu 138
Db 130043 CATTACTCTCAGATCGGGAATTTCCCTCTCTTTCAATGAGATAGCTCTTATTCCAATAA 130102
QY 139 AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
Db 130103 ACCAGATGAACCTCATGGATATCGCCAGATCTTCTCTCTTCTGCAACGAGAAAGAG 130162
QY 159 GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyAspAsnAsp 178
Db 130163 GAGGAGGACCCGACACAGCGGAAAGACAGCTCTCAATCAAGCTGTATATGACAATGAC 130222
QY 179 SerTyThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
Db 130223 TCTTGACCTGGACCACTTCTACACGAGAACGTTTATAACGGTTCATCAACAGCAGG 130282
QY 199 SerGlyTyrProGlyThrProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
Db 130283 AGTGGCTGGGTATACCTGGAAACACCTTGGCTTGGCAGCTCTTATGGTCACTTAGAT 130342
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
Db 130343 TGTGTGAAGTCTCTCGAACAATGGTGTGATGATGATGATGATGATGATGATGATGATGAT 130402
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
Db 130403 ACACCACTTTTCACTGCTGTAGTACGAGTATCGAGTGTGTGAGATACCTTTTAGAA 130462
QY 259 AlaGlyAlaSerProGlyLysSerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
Db 130463 GCTGGTGGCTGCTCTAGTGTAGTACATCTACAACTGTCTCTCTCTCTCTCTCTCTCT 130522
QY 279 ArgAspGlyAlaValAlaLeuLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
Db 130523 CGTGATGGGCTTGGCATCTTACAGGAGCTCTAGGAGCTCTAGGAGCTCTAGGAGCTCTAG 130582
QY 299 LysAlaLysLeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeu 318
Db 130583 AAGCGCAACTGCGAGTCTGGGATCAACATAGCTTCAATGTTCTGGCCCCCTCTATTG 130642
QY 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuLeuHisGlyAlaAspPro 338
Db 130643 GCTGCACTATAGTCTATCTTGTATGTTTCCGCTCTCTTGTCTCTATGGGCGAGATCT 130702
QY 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeu 358
Db 130703 GATTACAACGTATTGACACGAGGCTCTTAAGTCGAGTTCCACAGCTCGCACACTCTT 130762
QY 359 GluIleCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGly 378
Db 130763 GAAATCGCTCTCATCAATAATTGTGAGCCAGAGTACATACAGCTTTTAAATAGATTTTGA 130822
QY 379 AlaAsnIleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspLysGlyIle 398
Db 130823 GCGAATCTACTCTCCATCTCTACTCTTGGACCCCAACATCTCAAGATGATTAAGCATC 130882
QY 399 AlaLeuLeuGlnAlaArg-----405
Db 130883 AAGTTCTGCTACAAGCCGAGGTGAATAGTATGATCAGGCATCCTCATAGGCTTCTTTAT 130942
QY 405 -----405
Db 130943 TGGAGGTCTCAGGGTTCCAAATCATTAAGACCTTCCCATGGTGTAGCAAAATGGCTC 131002
QY 405 -----405
Db 131003 CATCATCTCTGCTGCTCTGAGAGAAAAGTTAAGATTTCAACAGATGAAGGCT 131062
QY 405 -----405
```

```
Db 131063 TACTAGATGAGGTGTGTAGAAAGCATTTATACAGTCTCTGGCCTTAGATCTGCCAATGA 131122
QY 405 -----405
Db 131123 CTCACCTCTTTTGTGATGCCCTGAGCTTCTCTTATAGGCCCTTGGTGTATGTTACCATGCAT 131182
QY 405 -----405
Db 131183 TTGGAGATGATTTGTGACATACTTAGACGGGAGAGGCATTATATAGCATGCCAGAT 131242
QY 405 -----405
Db 131243 AGATGACACTTTGGACAGAATAAGTACACAGAGAATCGATGCTCATATAATTTGTTTCTCAT 131302
QY 406 -----AlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeu 423
Db 131303 TTTTGCAGCTACTCTCACGGTCACTCTCTGCCAGGCCCTTAGTTATCCGAGATGCCCT 131362
QY 423 uCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSe 443
Db 131363 GTGCCGAGCAACACGACGACACAGCCATCGACCATGATATCCCCCTGTGTGATTAG 131422
QY 443 rTyrLeuLysHisGln 448
Db 131423 CTACCTCAAGCATCAA 131438

RESULT 10
LOCUS AL844570 240816 bp DNA linear VRT 11-FEB-2003
DEFINITION Zebrafish DNA sequence from clone DKEY-20D18, complete sequence.
ACCESSION AL844570
VERSION AL844570.9 GI:28300586
KEYWORDS HTG
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 240816)
Sehra, H.
Direct Submission
Submitted (11-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2003 this sequence version replaced gi:27652710.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
beginning 'Dr' were identified by the Recon repeat discovery system
(Zhirong Bao and Sean Eddy, submitted), and those beginning 'dr'
```

were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.Projects/D_rerio/fishmask.shtml
 DKEY-20D18 is from a Zebrafish BAC library
 VECTOR: pindisBAC-5

FEATURES

```

source
  1. 240816
  /organism="Danio rerio"
  /mol_type="genomic DNA"
  /db_xref="taxon:7955"
  /clone="pKEY-20D18"
  /clone_lib="DanioKey"

BASE COUNT 76749 a 43368 c 43554 g 77145 t
ORIGIN

Alignment Scores:
Pred. No.: 1,17e-70 Length: 240816
Score: 899.50 Matches: 178
Percent Similarity: 75.08% Conservative: 45
Best Local Similarity: 59.93% Mismatches: 61
Query Match: 38.39% Indels: 13
DB: 5 Gaps: 4

US-09-941-831A-20 (1-449) x AL844570 (1-240816)

QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
DB 18737 ATGAGTTTCATGGACATCTTAAGATCTTTCTCTTCAACCC---AAAGAGGATGAG 18793
QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyAspAsnAspSerTy 180
DB 18794 GAGGAAGACACCAACTCGAGTCACGAACTAAATCAAGCTGTCTTCACAGACGACGACAA 18853
QY 181 ThrLeuAspGlnLeuLeuArgGlnGluAgtTyLysArgPheIleAsnSerArgSerGly 200
DB 18854 CTCCTTACCTGAGCTGTGTCTCAGGAACAATATAGAAATGTCATCAACACCGCATGTG 18913
QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyGlyHisLeuSerCysLeu 220
DB 18914 TGGGGATCCCATCTCACTCCGCGACGGCGCTGCTCAGGGTCACCTACGCTGTCTG 18973
QY 221 GlnValLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240
DB 18974 GAGGTCCTCTGCTCAGCGGCGAGAGTCGACAGCGCTGATGTCAAAGCTCAACACCA 19033
QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
DB 19034 CTTTTCACAGCCGCTCTGCGAACACATAGATTGTGTGCTCTTTATTAAGAGCTGCG 19093
QY 261 AlaSerProGlyGlySerIleTyAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
DB 19094 GCAGACCTTANTGGAGGCCACATACACTGCTCCCTGTGTGACTGGCGCCAGAGAG 19153
QY 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
DB 19154 GGGGATGTGGACATCTCTGAAGAGGCTCTTCAATATATGGCGCGATGTAGACGTCAGGCC 19213
QY 301 LysLeuProValTrpAlaSerAniIleAlaSerCysSerGlyProLeuTyLeuAlaAla 320
DB 19214 AAAATGCCAGACTGGGCGCTTAAATGCCACAGCTTGGCGAGGACCACTGTACATTTCCAGCG 19273
QY 321 ValTyGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTy 340
DB 19274 GTGTATGGTCACTAGTGTGTTTCAATTTGTTCTGAGTTTGGAGCCCAACCCAGACTAC 19333
QY 341 AsnCysThrAspGlnGlyLeuAlaAlaArgValProArgProArgThrLeuLeuGluIle 360
DB 19334 AACTGCACTGAGGAAAAGATGCTGGCGCAGGATCAAGAGCGCTAACTGTTTGGAGATG 19393
QY 361 CysLeuHisIleAsnCysGluProGluTyIleGlnLeuLeuIleAspPheGlyAlaAsn 380
DB 19394 TGTCTGAGACATGGCTGTGGCGCTGGAATACATTCAGCTGCTCATAGACTTTGGAGCAAT 19453
QY 381 IleTyLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400

```

```

Db 19454 GTTATCTGCCCTTTAGTTGGAGATAAAACCCAGCTGAGAATGAACGAGTACTGCTT 19513
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
DB 19514 CTGCTTAAAGAGAGAGGTGAG-----TCTCAATTACAC----- 19546
QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIle 437
DB 19547 -----TTGTGTCAATCTTTAAG---ACCGGTTTCAACATTTTAGACATT 19588

```

```

RESULT 11
BC012399
LOCUS
DEFINITION
  Homo sapiens, Similar to hypothetical protein FLJ20126, clone
  MGC:8876 IMAGE:3864128, mRNA, complete cds.
ACCESSION
  BC012399.1 GI:15214550
VERSION
  BC012399.1
KEYWORDS
  MGC.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
  1 (bases 1 to 2653)
  Strausberg,R.
  Direct Submission
  Submitted (15-AUG-2001) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA

```

```

REMARK
  CONTACT: MGC help desk
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: DCTD/DP
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Baylor College of Medicine Human Genome
  Sequencing Center
  Center code: BCM-HGSC
  Web site: http://www.hgsc.bcm.tmc.edu/cdna/
  Contact: villalona@bcm.tmc.edu.
  Villalón, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
  A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
  Muzny, D.M., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 21 Row: j Column: 23
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7020020.

FEATURES

```

source
  1. 2653
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="MGC:8876 IMAGE:3864128"
  /tissue_type="Ovary, adenocarcinoma"
  /clone_lib="NIH_MGC_66"
  /lab_host="DH10B"
  /note="Vector: pCMV-SPORT6"
  69. 2183
  /codon_start=1
  /product="similar to hypothetical protein FLJ20126"
  /protein_id="AAH12399.1"
  /translation="MDHITVPKVENVKLVDRYVSKKPKANGILYLFATHLLIYVEASGAA
  RKTALHHTATVEKLPITSLGCLPLTRCKNFRVAHFVLDSDLVCHVEYIILKLSQ
  PALPDLAFSNPKSSKEMESGWLIDPISDFGRMGIPNENWTITDANRNVICST
  YPPELVKPSVTLGTGVSSEKSKERVPVLYLYKENNAALCRSQPSLGSYTRCVD
  DMLLEALISQTNPGSQFVVDTRPKLNANRAAGKYENEDNVTANIRFRFMGIENI
  HVRSLQKLELVCELKPTPTSEFSLGSLGRHKAIMDAGLFIITKAVKVEKASV
  LVHCSGDGWDRTAQVCVASILDDPFRPTFKGLMILIEKVISMGHFKFSRCCHLDGDS
  KEVSPIFTQFLDCIQLMEQPPCAFEFNFLLEIHDHVFVSCQFGNLCQCKREDLI

```

CDS

RYEKTHSVNPFVLPORPKDFRNPFLYKGTMYGVLPNSTVPIYNIQFWMYRPFKGLQ
PKQSMLESLLEIKKQAMLETDVHELEKLVKRDPEPPEICTCSQNLGNHLSGLPL
TNPLGPMGLINDLMTNGLTCTREGLRAQMDQVKSQGLADLHNCCEIVGSLRAINI
SGDVGISEANGISGDMCTPEATGFSKDLGICGMDISEATGISGNLISEARGFSGDM
GILGDTGISKASTKEADYSKHQ

BASE COUNT 749 a 566 c 645 g 693 t
ORIGIN

Alignment Scores:

Pred. No.: 1,46e-52 Length: 2653
Score: 675.50 Matches: 189
Percent Similarity: 50.95% Conservative: 52
Best Local Similarity: 39.96% Mismatches: 126
Query Match: 28.83% Indels: 107
DB: 9 Gaps: 19

US-09-941-831A-20 (1-449) x BC012399 (1-2653)

QY 1 MetArgGluSerGlyTrpLysLeuIleAspProIleSerAspPheGlyArgMetGlyIle 20
DB 432 ATGAGGGAAGTGGATGGAACTGATGACCAATATCAGACTTTGGCGGTATGGGAATA 491
QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
DB 492 CCCAACAGAACTGGACCATACAGATGCCAACAGAACTATCAGATATGCAGACCTAC 551
QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
DB 552 CTTCTCGAATAGTGTCTTAATCTGTACCTTGGGAACGGTGTGGAAAGTTCANAG 611
QY 61 PheArgSerLysGluArgValProValLeuSerTyrLysGluAsnAlaAla 80
DB 612 TTCAGAAATAAGACGTGCT 671
QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
DB 672 ATTTCGCGCTAGCCAGCGCTCTCTCTGATTTTACACTCGCTGTGTAGATGATGAGCTC 731
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
DB 732 TTGTTGGAGCCATTAGCCAAACACCCAGGAGCCAGTTATGTATGTATGTATGTATGTAT 791
QY 121 ArgProLysIle----- 124
DB 792 AGACCAAGTGAATGCCATGCCACACCGAGCAGCTGGGAGGGGTATGAAATGAAGAC 851
QY 125 -----TrpHisPheLeu-----ValLeuIleMetArgIle 134
DB 852 AACTATGCCAACATTCGCTTCAGATTTCATGGGCATTGAGAACATCCATGTAATGCGGAGC 911
QY 135 ValLeuGlnLeuAlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGln 154
DB 912 AGTCTCAG-----AACTCTTGGAAAGTTGTGAATGAAACT 950
QY 155 ProAspLysGluGluAspThrAspThrGluGluLysGlnAlaLeuAsn-----Gln 172
DB 951 CCACAACTAGTGAATTTCTAGCGCCCTGGAGAGCTCAGGGTGTAGACACATTAA 1010
QY 173 AlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLys 192
DB 1011 GCTATTATGATGCTGGAATTTTC---ATTACAAAGGCAGTGAAGGTAGAAAGGCCAGT 1067
QY 193 ArgPheIleAsnSerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAla 212
DB 1068 GTCTTAGTCCATTGTTCTGATGATGGGACCGCAGCAGCAAGCTGTCTCAGTGGCTAGC 1127
QY 213 -----SerTyrGlyHisLeuSerCysLeuGlnValLeuLeu----- 224
DB 1128 ATCTCTCTAGATCCATTTTATAGGACATTCAGAGGACTCATGCTTGTATAGAGGAA 1187
QY 225 -----AlaHisGlyAlaAspValAspSer 232
DB 1188 TGGATATCATCGGGCCACAAGTTTTTCCCAAGAGGTGTGGCCAC---CTCGATGGGACTCT 1244

QY 233 LeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCys 252
DB 1245 AAGAAGTG-----TCCCTATCTTCACCCAG-----TTCTAGACTGT 1283
QY 253 ValArgValLeuLeuGlu-AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys-- 271
DB 1284 ATCTGGCAATTAAATGAACAGTTTCCCTGTGCTTTGAGTTTAAATGAAACTTCCTGCTG 1343
QY 272 -----SerPro-----ValLeuThrAlaAla 279
DB 1344 GAGATTTCATGACCATGTTTCTCTCTGCCAGTTTGGAAACTTCCTTGTGTAATGCCAGAAG 1403
QY 279 GAspGlyAlaVal-----AlaIleLeuGlnGluLeu----- 290
DB 1404 GATCGGGAAGTCTAAGAGTCTATGAGAAACACATTTCTGTGTGGCTTCTTGTGGTTCAG 1463
QY 291 -----AspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrpAlaSerAs 308
DB 1464 AGAAACACAGACTTCAGGAACCTCTCTATATAAGGCTTCACCTATGTATGGGTACTCAAT 1523
QY 308 nIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysph 328
DB 1524 CCTAGTACTGTGC-----CCTACAACTTCAGTTTCTGTGTGGATGTATAACCGCT-- 1575
QY 328 eArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeu 348
DB 1576 -----TTGACAAAGGGCTGCAGCCCAAGCAGATATGCTA-GAGAGC-----CT 1618
QY 348 uAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHis----- 363
DB 1619 CTGGAAATTAAAGAACAGAGCAATGCTGCAGACAGATGTGCATGAACTAGAAAAGAA 1678
QY 364 -----HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAla 380
DB 1679 ACTAAAGTCCGTGATGAGCCACCAAGAGATCTGTACTCTCTCAATTAGGA---AA 1735
QY 380 nIleTyrLeuProSerLeuSerLeuAspLeuSer 392
DB 1736 CATATTATCCAGCATCTGGGAAGTCTTTGACCAAT 1772
RESULT 12
LOCUS AX406008 1345 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 423 from Patent WO0222660.
ACCESSION AX406008
VERSION AX406008.1 GI:21439455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Tang, Y. T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q. A., Ren, F.,
Xue, A. J., Yang, Y., Wehrman, T. and Drmanac, R. T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 423 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES
Location/Qualifiers
1..1345
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
92..1204
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35115.1"
/db_xref="GI:21439456"
/translation="MDHITVPKVENVKLVDRYVSKKPPANGILLYLTATHLIYVEASGAA
RRTWIAUHHIATVEKLPITSGCPLTRKFNVRVAFVLDSDLVCHVEVYLSLLKLSQ
PALPDLFYAFSNPKSKEMRESGKLIDPISDFRGMGIPNRNWTITDANRNYICST
YPIRIVVPKSVTLGTGVGSKRSKRPVILSVLYKENNAICRLSPLSEYTRCVD
DELLLEAISQTPGSMYVVDTRPKAVKVEKASVLVHCSGDGMDRTAQVCSVASILLD


```

303 QY -----ProValTrrAlaserAsnIleAlaser-----CysSerGly 314
624 Db ACTCTGATGTCAGCCTCGATCTCCCGGGCTCACCTCTCTGGTGTGCTGC----- 677
315 QY ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
678 Db CCCTTGTATACAGCGGCACCTACCAACACCTCCAGTGTCTCCGGCTGCTCTCTCTGGCT 737
335 QY GlyAlaAspProAspTyrAsnCys-----ThrAspGlnClyLeuLeuAlaArg 350
738 Db GCGCGAACCCCTGACTTCACTGCAATGCTCTGTCAACACACAGGATTCACAGG--- 794
351 QY ValProArgProArgThrLeuHisGlnCysLeuHisAsnCysGluProGluTyr 370
795 Db ---GGCTCCCTGGTGGCTGCTGATGCTGTCTGGCCACGGCTGTGAGCAGCCCTTC 851
371 QY IleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeu-----ProSerLeuSer 387
852 Db GTGAGCCTGCTGTAGAAATTTGAGCCAACTGAATCTAGTGAAGTGGGAATCGCTGGGC 911
388 QY LeuAspLeuThrSerGln-----AspAspLysGlyIleAlaLeuLeuGlnAla 404
912 Db CCAGAGTCGAGAGGAGAGAAAGTGGACCTTGAGCCTTCGAGCTTTTAAAGAGGCC 971
405 QY ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
972 Db AGAGGTGTTCCAGACACCTTGTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1031
425 QY GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyr 444
1032 Db AAA---CACCGGCTTCATCTGATCTCTCGTCTGCTCTGCCAGACCCCATAAAGATT 1088
445 QY LeuLysHisGln 448
1089 Db CTACTCCATGAG 1100

RESULT 15
AF155352
LOCUS
DEFINITION
Mus musculus ankyrin repeat-containing protein Asb-1 mRNA, partial
cds.
ACCESSION
AF155352
VERSION
AF155352.1 GI:5053021
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1008)
Kile, B.T., Viney, E.M., Willson, T.A., Brodnicki, T.C., Cancilla, M.R.,
Herlihy, A.S., Croker, B.A., Bacá, M., Nicola, N.A., Hilton, D.J. and
Alexander, W.S.
Cloning and characterization of the genes encoding the ankyrin
repeat and SOCS box-containing proteins Asb-1, Asb-2, Asb-3 and
Asb-4
JOURNAL
Gene 258 (1-2), 31-41 (2000)
MEDLINE
20564172
PUBMED
11111040
REFERENCE
2 (bases 1 to 1008)
Kile, B.T., Viney, E.M. and Willson, T.A.
Direct Submission
AUTHORS
Submitted (01-JUN-1999) Cancer and Hematology, The Walter & Eliza
Hall Institute of Medical Research, Royal Parade, Parkville, VIC
3052, Australia
LOCATION/Qualifiers
1..1008
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
1..>1008
/note="contains SOCS box"

FEATURES
source
CDS

```

```

/codon_start=1
/product="ankyrin repeat-containing protein Asb-1"
/protein_id="AAD38808.1"
/db_xref="GI:5053022"
/translation="MAEGGTGDPGRAGPAGPNLKEWLRFQFDHPLEHCDTDLRKH
AAYVDLQTLRLNLIQESYRINEKSVCCGWLPTCLRIATAATAGHCNVDPLIRKH
AEVLVDVKGQTVALLYAVVNGHLESTELLLEAGADPNGRHRRSTPVVHARVGRDDI
LKALIRYAGDVAVNNHLLTPDRPPPSRLTSLVCLPYSIAAHNLQCFRLLLQAGAN
PDENCNGPVNTQEFYRSGPGCVMDAVLRHGCEAAVSLVBFGANLNLKWSLSQPEA
RGRKMDPEALQVPKEARSIPRTLILLSLCEVAVRRALGKYLHLVSLPLPDPFKFLL
YE"
BASE COUNT      210 a      272 c      306 g      220 t
ORIGIN
Alignment Scores:
Pred. No.:      4,2e-35      Length:      1008
Score:          481.00      Matches:    114
Percent Similarity: 58.0%      Conservative: 59
Best local Similarity: 38.26%      Mismatches: 103
Query Match:    20.53%      Indels:     22
DB:             10          Gaps:        8
US-09-941-831A-20 (1-449) x AF155352 (1-1008)
QY 170 LeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuArgGlnGlu 189
Db 124 CTCCATGATGACGACCTATGAGGGACCTCCAGACCTCAGGAACCTACTGCAAGAGGAG 183
QY 190 ArgTyrLysArgPheIleAsnSerArgSerGlyTyr-----GlyValProGlyThr 206
Db 184 AGCTACCGGACGCGCATCAATGAGAAGTCTGTCTGCTGCTGGCTGCTCCCTGCACA 243
QY 207 ProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHis 226
Db 244 CCACCTGAGGATCGCAGCCACTGCAGGCCATGGGAATCTGTGTGAGCTTCCTACGCAA 303
QY 227 GlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSer 246
Db 304 GGGGCGAGGTGGACTGTGTGATGTCAAGGGGAGAGCTGCCCTGTATGTGCTGTAGTG 363
QY 247 HisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySer 266
Db 364 AACGGGCACTTGGAGACACTGAGATCTCTTTTGGAGAGCTGTGTGATCCCAACGCGCAG 423
QY 267 IleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeu 286
Db 424 CGGCACCCAGCAGCAGCTCTGTGTACCATGCTCTCTGTGTGGTAGGAGCAGCATCTCG 483
QY 287 GlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu----- 302
Db 484 AAGGCTCTTATCAGGTATGGGCGCAGATGTGTATGTCAACCATCATCTGACTCTGACACC 543
QY 303 ---ProValTrrAlaserAsnIleAlaser-----CysSerGlyProLeuTyrLeu 318
Db 544 CGGCCCTCTTTTACGCGCGCTAACCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 597
QY 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 598 AGTGTGCTTACCAATACCTTTCAGTGTTCAGGCTGCTCTTCAGGCTGGGGCAATCTCT 657
QY 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArg-----ProArgThr 356
Db 658 GACTTCAATTTGAATGGCCCTGTCAACACCCAGGAGTCTTACAGGGATGCCCTCTGGGTGT 717
QY 357 LeuLeuGluIleCysLeuHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAsp 376
Db 718 GTCATGGATGCTGCTCTGCGCCATGGCTGTGAAGCAGCCCTTCGTGAGTCTGTGTGTAGAG 777
QY 377 PheGlyAlaAsnIleTyrLeu-----ProSerLeuSerLeuAspLeuThrSerGln 393
Db 778 TTTGGAGCCCACTGAACTGTGAAGTGGGAATCCCTGGGCCAGAGGAGCAAGAGGCAGA 837
QY 394 -----AspAspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSer 410

```

```

Db      838 AGAAGATGATCCGAGGCTTGCAGGTCTTTAAAGAGGCCAGAGTATTCGAGGACC 897
Qy      411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db      898 TTGCTGAGTTGTGCCGGGTGGCTGTGAGAAGAGCTCTTGCAAAATAC--CGACTGCAT 954
Qy      431 AlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyrLeuLysHisGln 448
Db      955 CTGTTCCCTCGCTGCCGCTGCCAGACCCCAAAAGAGTCTTTGCTTTATGAG 1008

```

Search completed: December 1, 2003, 09:17:07
 Job time : 4616 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:38:12 ; Search time 353 Seconds
(without alignments)
3433.563 Million cell updates/sec

Title: US-09-941-831A-20

Perfect score: 2343

Sequence: 1 MRESGWKLDIPDFGRMG.....QAINQLDIPPMISLYLKHL 449

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_p2n.model -DEV=xl
-Q=/cgn2_1/USProoool_p/US09941831/runat_01122003_073526_7819/app_query.fasta_1.647
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptn -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09941831 -CGEN 1 1 396 -runat_01122003_073526_7819 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2343	100.0	1878	22	AAH46857	Human serine/threose
2	2343	100.0	1878	22	AAH78730	Human HIBC989 seri
3	2343	100.0	1878	22	AAS33372	DNA encoding human
c	2331	99.5	2530	22	AAS33264	DNA encoding human
5	1690	72.1	1222	24	ABK11107	DNA encoding human
6	1680	71.7	1127	24	ABZ11533	Human polynucleoti
7	1629	69.5	1069	24	ABK11106	DNA encoding human
c	672	28.7	8214	23	AAS77257	DNA encoding novel
9	642.5	27.4	1345	24	ABN60012	Novel human coding
c	499	21.3	1266	22	AAS33071	DNA encoding human
10	499	21.3	1433	23	ABV23231	Human prostate exp
11	499	21.3	1433	23	ABV29074	Human prostate exp
12	499	21.3	1509	25	AAD49617	Human cytoskeleton
13	499	21.3	1540	22	AAS33289	DNA encoding human
14	499	21.3	1540	22	AAS33289	Mus musculus SOCS7
15	480	20.5	2019	19	AAV38672	Human phosphatase
16	424.5	18.1	2353	22	AAF63581	Human cDNA SEQ ID
17	394	16.8	2482	24	ABQ93336	Drosophila melanog
18	378.5	16.2	2786	23	ABL06295	Drosophila melanog
19	378.5	16.2	3047	23	ABL20003	Drosophila melanog
20	378.5	16.2	3169	23	ABL17297	Drosophila melanog
c	366	15.6	408	23	ABV61004	Human prostate exp
c	335	14.3	8397	23	ABL20002	Drosophila melanog
c	335	14.3	9648	23	ABL06294	Drosophila melanog
c	335	14.3	9648	23	ABL17296	Drosophila melanog
25	297	12.7	1966	22	AAH14752	Human cDNA sequenc
26	293.5	12.5	545	23	AAS73111	DNA encoding novel
27	287	12.2	3382	24	ABK33507	Human cDNA differe
28	281	12.0	2710	22	AAK32009	Human polynucleoti
29	281	12.0	4330	24	AAD40741	Human kinase and p
30	275.5	11.8	2575	23	AAS73882	DNA encoding novel
31	273	11.7	2346	22	AAK32993	Human polynucleoti
32	270.5	11.5	1622	21	AAK36830	DNA encoding a su
33	264.5	11.3	2764	22	AAH17868	Human cDNA sequenc
34	264	11.3	3411	18	AAT91998	Human tyrosine pho
35	261.5	11.2	1206	22	AAI58158	Human polynucleoti
36	261.5	11.2	3398	24	AAI71054	Benign prostatic h
37	257.5	11.0	1024	24	ABQ54306	Human ovarian anti
38	252.5	10.8	1869	22	AAH13777	Human cDNA sequenc
39	252.5	10.8	2263	21	AAZ36833	DNA encoding a su
40	246	10.5	802	22	AAH07276	Human cDNA clone (
41	246	10.5	1538	24	ABA93131	Human C/SKARP-1 pr
42	244.5	10.4	401	21	AAA77749	cDNA encoding huma
43	244.5	10.4	401	22	AAI28487	Colon tumour relat
44	244.5	10.4	401	25	ABZ32673	Human colon tumour
45	241	10.3	1565	22	ABA09163	Human cytokine sig

ALIGNMENTS

RESULT 1
AAH46857

ID AAH46857 standard; cDNA; 1878 BP.

XX AAH46857;

XX 25-SEP-2001 (first entry)

XX Human serine/threonine phosphatase encoding cDNA (clone ID HIBC989).

KW Serine/threonine phosphatase; nootropic; neuroprotective; cytostatic;
KW dermatological; immunosuppressive; antiinflammatory; antibacterial; ss;
KW anti-HIV; antiparkinsonian; antischistosomal; antianemic; antirheumatic;
KW antirheumatic; virucide; hepatotropic; cerebroprotective; vulnerary;
KW antiinflammatory; nephrotropic; gene therapy; vaccine.
XX Homo sapiens.
XX OS

XX	Key	Location/Qualifiers	Db	427	CCAAACAGAACTGGACCATACAGATGCCAACAGAAATATGAGATATGCAGCACCTAC	486
FH	CDS	367..1716	QY	41	ProProGluIleValValProlySerValThrLeuGlyThrValValGlySerSerLys	60
FT		/*tag= a	Db	487	CCTCTGAAATAGTGGTTCTTAATCTGTATCTTGGGAACGGTGGTTTGAAGTTCAAAG	546
PN	WO200155388-A1.		QY	61	PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAla	80
XX			Db	547	TTCAAGAGTAAGAAGCGTCTCTGTCTCTCTACCTCTACAAGAGAACAAATGCTGCC	606
PD	02-AUG-2001.		QY	81	IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu	100
PF	17-JAN-2001; 2001WO-US01395.		Db	607	ATTTGCGCTGTAGCCAGCCTCTCTCTGATTTTACACTCGCTGTGTAGATGATGAGCTC	666
XX	31-JAN-2000; 2000US-0179065.		QY	101	LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr	120
PR	04-FEB-2000; 2000US-0180628.		Db	667	TTGTTGGAGGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTGTAGACACA	726
PR	02-MAR-2000; 2000US-0186350.		QY	121	ArgProLysIleTyrPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys	140
XX	(HUMA-) HUMAN GENOME SCI INC.		Db	727	AGACCAAGATCTGGCATTTCTGTGTCTCATATAGAGATAGTTCTCCAATTAGCCAAAG	786
PI	Rosen CA, Barash SC, Ruben SM;		QY	141	MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu	160
XX	WPI; 2001-476208/51.		Db	787	ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCTGCAGCCCGCACAAGGAGGAG	846
DR	P-PSDB; AA865477.		QY	161	AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr	180
XX			Db	847	GACACTGCACAGAGAGAGAGAGAGCTCTCTCAATCAAGCAGTGTATGACAAACACTCTAT	906
XX			QY	181	ThrLeuAsnGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly	200
CC			Db	907	ACTTTGGACCACTTTTGGCCAGAGCGTTACAACAGTTTCATCAACAGCAGGAGTGGC	966
CC			QY	201	TyrGlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeu	220
CC			Db	967	TGGGTGTCTCTGGGACACCTTGGCTGTGGCTTCTTATGGCCACTTGGCTGTTTG	1026
CC			QY	221	GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro	240
CC			Db	1027	CAAGTCTCTTAGCCCATGGTCTGTATGTGACAGCTTGGATGTCAAGSCACAGAGCCA	1086
CC			QY	241	LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly	260
CC			Db	1087	CTTTTCACCTGCTGTGACATCGCCATCTGAGTGTGTACGTGTGCTTTTGAAGCTGT	1146
CC			QY	261	AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp	280
CC			Db	1147	GCCTCTCTGTGTGTAGCATCTACAACAACTGTCTCCCGTCTCACAGCTGCCGTGAT	1206
CC			QY	281	GlyAlaValAlaIleLeuGlnLeuAspHisGlyAlaGluAlaAsnValLysAla	300
CC			Db	1207	GGTGTGTGTGTATCTCTGCAGGAGCTCTTAGACCATGTGTGACAGAGCCCAACGCTCA	1266
CC			QY	301	LysLeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla	320
CC			Db	1267	AAACTACAGCTCTGGGCATCAAAACATAGCTTCTGTGTCTGGCCCTCTTATTTGGCCGA	1326
CC			QY	321	ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr	340
CC			Db	1327	GTCTACGGGCACCTGGAGCTGTTTCCGCTGTCTTTTGTCTCCACGGGCGACCTGACTAC	1386
CC			QY	341	AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuLys	360
CC			Db	1387	AACTGCACCTGACACAGGCGCTATTGGCTGTGTGTCTCCCAAGACCCCGACCTCTTGAATC	1446
CC			QY	361	CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAsnPheGlyAlaAsn	380
CC			Db	1447	TGCTCCCATCAATATGTGAGCCAGAGTATATCCAGCTGTATTAATGATTTTGTGCTAAT	1506
CC			QY	381	IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu	400
CC			Db	1507	ATCTACCTTCCATCTCTCTCCCTTGCCTGACCTCACAAGATGATAAAGGCAATGCTATTG	1566

Novel proteins of serine/threonine phosphatase family, useful for diagnosing, treating, preventing and/or prognosing disorders related to the proteins, including cancer, immune response and neuronal disorders

Claim 1; Page 426; 441pp; English.

The invention provides serine/threonine phosphatase family polypeptides and polynucleotides encoding them. The polypeptides can be expressed by standard recombinant methodology. The polypeptides, polynucleotides and antibodies are useful for diagnosis, prognosis, prevention and treatment of neurodegenerative disorders, immune system disorders, autoimmune diseases, allergic reactions, infectious diseases, hyperproliferative disorders, renal disorders, cardiovascular disorders, cerebrovascular disorders, respiratory disorders, endocrine disorders, gastrointestinal disorders and also muscular, reproductive disorders (see AAH46846 for a detailed description of the diseases that can be treated). They are also useful as immune system enhancers, immunosuppressive agents, stimulators of B-cell responsiveness to pathogens, activators of T-cells, to induce higher affinity antibodies, as an agent to increase serum immunoglobulin concentrations, to accelerate recovery of immunocompromised individuals, to boost immunoresponsiveness in aged populations and/or neonates, as a regulator of antigen presentation, as a means to induce tumour proliferation in pathologies e.g. AIDS, and/or common variable immunodeficiency. The polypeptides and polynucleotides are useful to prevent skin aging, for preventing hair loss, to stimulate growth and differentiation of hemopoietic cells and bone marrow cells, for supporting cell culture of primary tissues, to modulate mammalian characteristics such as body weight, height, eye color, hair color and skin, to modulate mammalian metabolism to change a mammal's mental or physical state, and as food additive or preservative. The present sequence represents a human serine/threonine phosphatase encoding cDNA.

SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Alignment Scores:

Pred. No.:	2,616-230	Length:	1878
Score:	2343.00	Matches:	449
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-941-831A-20 (1-449) x AAH46857 (1-1878)

QY	1	MetArgGluSerClyTyrLysLeuIleAspProIleSerAspPheGlyArgMetGlyIle	20
Db	367	ATGAGGGAAGTGGATGGAAACTGATTCACCCCAATATCAGACTTTCGGCGTATGGGAATA	426
QY	21	ProAsnArgAsnTyrThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr	40

QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
 DB 1567 CTGCTACAGCCGCGAGCACTCCAGGTCACCTTATACAGGTCGGTTAGCTCGCGC 1626
 QY 421 ArgAlaLeuGlnAlaGlnProGlnAlaAlaAsnGlnLeuAspLeuProMet 440
 DB 1627 AGAGCCTTGCGCAGGCTGCGCAGGACCAAGCCATCACAGCTGGATATCTCCCATG 1686
 QY 441 LeuLeuSerTyrLeuLeuHisGlnLeu 449
 DB 1687 TTGATTAGCTACTAAACACCAACTG 1713
 RESULT 2
 AAH78730
 ID AAH78730 standard; cDNA; 1878 BP.
 XX
 AC AAH78730;
 XX
 DT 03-DEC-2001 (first entry)
 XX
 DE Human HIBCJ89 serine/threonine phosphatase cDNA sequence.
 XX
 KW Human; HIBCJ89; ss; serine/threonine phosphatase; PSpase; vaccine;
 KW gene therapy; PSpase expression; PSpase expression; PSpase modulation;
 KW immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome;
 KW Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis;
 KW inflammation; Crohn's disease; inflammatory bowel disease; appendicitis;
 KW rheumatoid arthritis; cellular proliferative disorder; lymphoma;
 KW lung cancer; intestinal cancer; cardiovascular disorder; aneurysm;
 KW Scimitar syndrome; Ebstein's anomaly.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 367..1716
 FT /*tag= a
 FT /product= "HIBCJ89 PSpase protein"
 XX
 PN WO200164703-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06256.
 XX
 PR 02-MAR-2000; 2000US-0186350.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ebner R, Ruben SM;
 XX
 DR WPI; 2001-530113/58.
 DR P-PSDB; AAG77803.
 XX
 PT Nucleic acids encoding serine/threonine phosphatase polypeptides,
 PT useful for preventing, diagnosing and/or treating, e.g. Crohn's
 PT disease, lung cancer and Scimitar syndrome -
 XX
 PS Claim 1; Page 308; 335pp; English.
 XX
 CC The present sequence represents the specifically claimed human cDNA clone
 CC HIBCJ89 which shares homology with members of the serine/threonine family
 CC of phosphatases. The invention comprises novel human serine/threonine
 CC phosphatase (PSpase) polypeptides and polynucleotides which may be used
 CC in the prevention (vaccine), diagnosis and treatment (gene therapy) of
 CC diseases associated with inappropriate PSpase expression. The PSpase
 CC polynucleotides of the invention may be used as DNA probes to detect and
 CC quantitate the presence of similar nucleic acids in samples. The PSpase
 CC polypeptides may be used as antigens in the production of antibodies
 CC against the PSpase polypeptides and in assays to identify modulators of
 CC PSpase expression and activity. The anti-PSpase antibodies and
 CC antagonists may also be used to down regulate expression and activity,
 CC the anti-PSpase antibodies may also be used as diagnostic agents for

CC detecting the presence of PSpase polypeptides in samples. Disorders that
 CC may be prevented, diagnosed and/or treated by the invention are:
 CC immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome,
 CC Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple
 CC sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory
 CC bowel disease, appendicitis and rheumatoid arthritis); cellular
 CC proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and
 CC cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and
 CC aneurysm).
 XX
 SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Alignment Scores:

Pred. NO.: 2,61e-230 Length: 1878
 Score: 2343.00 Matches: 449
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-941-831A-20 (1-449) x AAH78730 (1-1878)

QY 1 MetArgGluSerGlyTyrPlysLeuLeuAspProIleSerAspPheGlyArgMetGlyIle 20
 DB 367 ATGAGGAAAGTGGATGAAACTGATTGACCAATATCAGACTTTGGCGTATGGAATA 426
 QY 21 ProAsnArgAsnTyrThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
 DB 427 CCCAACACGAAATGACCATACAGATGCCAACGAACTATGAGATATGACGACCTAC 486
 QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
 DB 487 CCTCTGAAATAGTGGTCTTAATCTGTTACCTGGGAACGCTGTTGGAAGTTCAAG 546
 QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
 DB 547 TTCAGAGTAAAGAACGTGCTCCTGTCTCTCTACCTCTACAAAGAGAACAAATGCTGCC 606
 QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
 DB 607 ATTGCCCGCTGTAGCAGCCTCTCTGGATTTTACCTCGCTGTGTAGATGATGAGCTC 666
 QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
 DB 667 TTGTTGGAGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA 726
 QY 121 ArgProLysIleTyrPhePheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
 DB 727 AGACCAAAAGATCTGGCATTTCTTGTGCTCATATGAGAATAGTTCTCCAATTAGCCAAG 786
 QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
 DB 787 ATGAACCTCATGGACATCACCAGATCTTCTCCCTCGCAGCCCGACAGAGGAGGAG 846
 QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
 DB 847 GACATGTACACAGAGAGAGCAGGCTCTCAATCAAGCAGTGTGTATGACAACTCTTAT 906
 QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
 DB 907 ACTTTGGACCAGCTTTTGGCCAGGAGCGTTTACAACAGTTTTCATCAACAGCAGGAGTGC 966
 QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeu 220
 DB 967 TGGGGTGTCTCTGGACACCTTGGCGTGGCTGCTTCTTATGGCCACTGTGAGCTGTTTG 1026
 QY 221 GlnValLeuLeuAlaHisGlyValAspValAspSerLeuAspValLysAlaGlnThrPro 240
 DB 1027 CAAGTCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAAGCAGCAGACGCCA 1086
 QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuAlaGly 260
 DB 1087 CTTTTCACGTGTGTCAGTCATGCCCATCTGGACTGTGTACGTGTGCTTTTGGAGCTGT 1146


```
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
Db 667 TTGTTGGAGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTCTAGACACA 726
QY 121 ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
Db 727 AGACCAAGATCTGGCATTTCTTGGCTCATTAATGAGAATAGTTCTCCCAATTAGCCAA 786
QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
Db 787 ATGAACCTCATGGACATACACAGATCTTCTCTCTCGACCCGACAGAGAGAGGAG 846
QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValIlyAspAsnAspSerTyr 180
Db 847 GACACTGCACAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTAT 906
QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
Db 907 ACTTTGGACGAGCTTTTGGCCGAGGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGC 966
QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeu 220
Db 967 TGGGGTGTCTCGGACACCTTGGCTTGGCTGTCTTCTATGACCATTTGAGCTGTTTG 1026
QY 221 GlnValLeuAlaHisGlyAlaAspValAspSerLeuAspValIlyAspAlaGlnThrPro 240
Db 1027 CAAGTCTCTTAGCCCATGCTGATGTTGACAGCTTGATGTCAAGGCACAGACGCCA 1086
QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
Db 1087 CTTTTCACCTGCTCAGTCATGCCATCTGGACATCTGTGTACGTCTGCTTGTGGAAGCTGT 1146
QY 261 AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
Db 1147 GCCTCTCTCGGTGGTAGCATCTACAACTGTCTTCCCGCTGTACAGCTGCCCGTGTAT 1206
QY 281 GlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValIlyAla 300
Db 1207 GGTGCTGTGCTATCTCGCAGGAGCTCTTAGACCATGTTGACAGGCCAACGTCAAAGCT 1266
QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320
Db 1267 AAACCTACCACTCTGGGCATCAACATAGCTTCATGTTCTGGCCCCCTCTATTGTCGCGCA 1326
QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
Db 1327 GTCTACGGGCACCTGGACCTGTTCGCCCTGCTTTTGTCTCCAGGGGACACCTGACTAC 1386
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
Db 1387 AACTGCACTGACGAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTCTTGAATC 1446
QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
Db 1447 TGCCTCCATCAATAATTGTGAGCAGAGTATATCCAGCTGTTAATCATTTTGGTGTAAAT 1506
QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
Db 1507 ATCTACTCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATGAAGGCATTGCAATG 1566
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
Db 1567 CTGCTACAGCCGCGAGCCACTCCAGGTCATCTTATCATCAGGTCGTTTAGTGTCCGC 1626
QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMet 440
Db 1627 AGAGCCTTGTGCCAGGCTGCCAGGACCAAGCCATCAACAGCTGGATATCTCTCCCATG 1686
QY 441 LeuIleSerTyrLeuLysHisGlnLeu 449
Db 1687 TTGATTAGCTACCTAAACCAACCACTG 1713
```

RESULT 4

```
AAS33264/c
ID AAS33264 standard; cDNA; 2530 BP.
XX AC AAS33264;
XX DT 04-DEC-2001 (first entry)
XX DE DNA encoding human secreted protein, Seq ID No 223.
XX KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cystostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX OS Homo sapiens.
XX PN WO200155326-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01347.
XX PR 31-JAN-2000; 2000US-0179065.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI: 2001-451931/48.
XX DX P-PSDB; AAU20555.
XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing
XX PT or treating medical conditions -
XX PS Claim 1; SEQ ID No 223; 753pp; English.
XX CC The invention relates to novel isolated nucleic acid molecules (I)
XX CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
XX CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
XX CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
XX CC the prevention, treatment and diagnosis of diseases associated with
XX CC inappropriate expression of secreted proteins. (I) and complementary
XX CC sequences may also be used as DNA probes in diagnostic assays (e.g.
XX CC polymerase chain reactions (PCR)) to detect and quantitate the presence
XX CC of similar nucleic acid sequences in samples, and so which patients may
XX CC be in need of restorative therapy. (II) may also be used as antigens in
XX CC the production of antibodies and in assays to identify modulators
XX CC (agonists and antagonists) of the expression and activity of the secreted
XX CC proteins. The anti-(II) antibodies and antagonists may also be used to
XX CC down regulate expression and activity of (II). The anti-(II) antibodies
XX CC may also be used as diagnostic agents for detecting the presence of (II)
XX CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
XX CC disorders include for example: immune/autoimmune diseases (e.g. HIV
XX CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
XX CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
XX CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
XX CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
XX CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
XX CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
XX CC angina and thrombosis), infections caused by bacteria, viruses and
XX CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
XX CC agonists, antagonists and antibodies can also be used to promote wound
XX CC healing, maintain organs before transplantation, and support cell culture
XX CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
XX CC coding sequences, PCR primers, and related sequences of the invention.
XX CC Note: the sequence data for this patent did not appear in the printed
XX CC specification but was obtained in electronic format directly from WIPO
XX CC at: ftp.wipo.int/pub/published_pct_sequences.
XX CC Sequence 2530 BP; 630 A; 602 C; 675 G; 614 T; 9 other;
```

Alignment Scores: 6.9e-229 Length: 2530
 Pred. No.: 2331.00 Matches: 447
 Score: 2331.00
 Percent Similarity: 99.55% Conservative: 0
 Best Local Similarity: 99.55% Mismatches: 2
 Query Match: 99.49% Indels: 0
 DB: 22 Gaps: 0

US-09-941-831A-20 (1-449) x AAS33264 (1-2530)

QY 1 MetArgGluSerGlyTrpLysLeuLeuAspProIleSerAspPheGlyArgMetGlyIle 20
 DB 1575 ATGAGGGAAGTGGATGGAAATGATGACCAATATCAKATTTGGCGGTATGGGAATA 1516
 QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
 DB 1515 CCCAAGAGAACTGGACCAATAMCAGATGCCAAGCAAGCAATATGAGATATGCACACCTAC 1456
 QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
 DB 1455 CCTCTCGAATAGTGGTTCCTAATCTGTACCTTGGGAACGGTGGTGGAACTTCAAG 1396
 QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
 DB 1395 TTCAGAAAGTAAAGAACGTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1336
 QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
 DB 1335 ATTTGGCGCTGTAGCCAGCGCTCTCTGGATTTTACACTCGCTGTGTAGATGATGAGCTC 1276
 QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
 DB 1275 TTGTTGGAGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGATGTTGTAGACACA 1216
 QY 121 ArgProLysIlePrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
 DB 1215 AGACCAAGATCTGGCATTTCTCTGTCTCATATAATGAGAATAGTCTCTCAATTTAGCCAAG 1156
 QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
 DB 1155 ATGAACCTCATGGACATCAACCAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1096
 QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
 DB 1095 GACACTGACACAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACCAACGACTCTCTAT 1036
 QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
 DB 1035 ACTTTGGACACGCTTTTGGCCAGGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGC 976
 QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeu 220
 DB 975 TGGGGTGTCTCTGGGACACCTTGGCGCTTGGCTCTCTTATGGCCACTTGAGCTGTTTG 916
 QY 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240
 DB 915 CAACTCTCTTAGCCCATGGT 856
 QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
 DB 855 CTTTTTCACTGTCTAGTCAATGCCATCTGGACTGTGTAGTGTGTGTGTGTGTGTGTGTGT 796
 QY 261 AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
 DB 795 GCCTCTCTCTGGTGTAGTACTATACAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 736
 QY 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
 DB 735 GGTGCTGTGTCTATCTCTGAGGAGCTCTTAGACCATGTGTGAGAGCCCAACGTCAAGCT 676
 QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320

DB 675 AAACCTACCAGTCTGGGCATCAACATAGCTTCTGTCGCCCCCTCTATTTTGGCCGCA 616
 QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
 DB 615 GHTACGGGCACTGGACHTTTTCCGCTGCTTTTGGCTCCAGGGGACACCTGACTAC 556
 QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
 DB 555 AACTGCACCTGACCGGCTATTGGCTCGTCTCCCAAGACCCCGCACCTCTCTTGAATC 496
 QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsn 380
 DB 495 TGCCTCCATCATATAATTGTGAGCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTAA 436
 QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
 DB 435 ATCTACCTTCCATCTCTCTCCCTTGACCTGAGCTCACAAGATGATAAAGCATTTGATTG 376
 QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
 DB 375 CTGCTACAGGCGCGAGCCACTCCAGGTCACCTTCTATCACAGTCCGTTTGTGTCGCGC 316
 QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMet 440
 DB 315 AGAGCCTTTGTGCGAGCTGGCCAGCCACCAAGCCATCAACAGCTGGATATTCCTCCCATG 256
 QY 441 LeuIleSerTyrLeuLysHisGlnLeu 449
 DB 255 TTGATTAGTACTACCTAAACACCAACTG 229

RESULT 5
 ABK11107
 ID ABK11107 standard; DNA; 1222 BP.
 XX
 AC ABK11107;
 XX
 DT
 XX
 XX 05-JUN-2002 (first entry)
 DE
 XX DNA encoding human NOV5b protein, homologue of ankyrin-repeat proteins.
 KW Human; NOVX-associated disorder; developmental disorder; blood disorder;
 KW endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
 KW respiratory disorder; inflammatory disorder; reproductive disorder;
 KW neurodegenerative disorder; autoimmune disorder; infectious disease;
 KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
 KW metabolic pathway modulation; NOV5b; gene; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 CDS 6..1088
 FT /*tag= a
 FT /product= "NOV5b"
 XX
 PN W0200206329-A2.
 XX
 PD 24-JAN-2002.
 XX
 PF 18-JUL-2001; 2001WO-US22709.
 XX
 PR 18-JUL-2000; 2000US-218870P.
 PR 18-JUL-2000; 2000US-218875P.
 PR 18-JUL-2000; 2000US-218901P.
 PR 24-JUL-2000; 2000US-220273P.
 PR 26-JUL-2000; 2000US-220912P.
 PR 27-JUL-2000; 2000US-221233P.
 PR 28-JUL-2000; 2000US-221650P.
 XX
 PA (CURA-) CUPAGEN CORP.
 XX
 PI Rastelli L, Shinkets RA, Zerhusen B, Malyankar UM, Padigaru M;
 XX WPI; 2002-179781/23.
 DR

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders -
XX
PS Claim 1; SEQ ID NO 415; 1012pp + Sequence Listing; English.
XX

CC The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1127 BP; 244 A; 329 C; 281 G; 273 T; 0 other;

Alignment Scores:
Pred. No.: 1 49e-162 Length: 1127
Score: 1680.00 Matches: 324
Percent Similarity: 99.69% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 71.70% Indels: 0
DB: 24 Gaps: 0

US-09-941-831A-20 (1-449) x ABZ11533 (1-1127)

QY 124 ileftrPhisPheLeuValLeuMetArgileValLeuGlnLeuAlaLysMetAsnLeu 143
DB 135 ATCTGGCATTTCTCTGTCTATAATGAGATAGTTCTCCATTATGCCAAGATGACCTC 194
QY 144 MetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAspThrAsp 163
DB 195 ATGGACATCACCAGATCTTCTCCCTCTCGAGCCGACAAAGAGAGAGGACACTGAC 254
QY 164 ThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAsp 183
DB 255 ACAGAGAGAGACAGGCTCTCATCAAGCAGGTGTATGCAACACACTCTTATCTTTGGAC 314
QY 184 GlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrpGlyVal 203
DB 315 CAGCTTTTGGCCAGAGCGTTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGGGTGT 374
QY 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
DB 375 CTGGGACACCTTGGCTGGCTGTCTTCTATGGCCACTTGACGTGTTTGGAAAGTCCTC 434
QY 224 LeuAlaHisGlyValAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThr 243
DB 435 TTAGCCCATGGTGTGATGTGACCTTGGATGTGACGACACAGCCACTTTTCACT 494
QY 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
DB 495 GCTGTCTAGTCATGGCCATCTGGACTGTGTACGTGTCTTTTGGAACTGGTGCCTCTCT 554
QY 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
DB 555 GGTGTGTAGCATCTACAAACACTGTTCTCCGTGCTCACAGCTGCCGTGATGTGTGT 614
QY 284 AlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuPro 303
DB 615 GCTATCTCAGAGAGCTCTTAGACCATGGTGGAGGCCAAGCTCAAGCTTAACCTACCA 674
QY 304 ValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323

DB 675 GTCTGGCATCAACATAGCTTTTCATGTTCTGGCCCCCTCTATTGCGCCGAGCTACGGG 734
QY 324 HisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThr 343
DB 735 CACCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGACAGCCCTGACTACAACTGCAC 794
QY 344 AspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHis 363
DB 795 GACCAGGGCTATTGGCTGTGTCCCAAGACCCCGACCTCTCTTGAATCTGCCTCCAT 854
QY 364 HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeu 383
DB 855 CATAATTGTAGCCAGAGATATATCCAGCTGTTTAATCGATTTTGGTGTCTAATATCTACCT 914
QY 384 ProSerLeuSerLeuAspIleThrSerGlnAspAspLysGlyIleAlaLeuLeuGln 403
DB 915 CCATCTCTCTCCCTTGACCTGACCTCACAAAGATATAAAGGCATTGCATTGCTGTACAG 974
QY 404 AlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeu 423
DB 975 GCCCGAGCCACTCCACGGTCACTTCTATCACAGGTCCGTTAGTCGTCGCGAGAGCTTG 1034
QY 424 CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSer 443
DB 1035 TGCCAGGCTGGCCAGCCACAGCCATCAACCAGCTGGATATCTCTCCCATGTTGATTAA 1094
QY 444 TyrLeuLysHisGlnLeu 449
DB 1095 TACCTAAAACACCACCTG 1112
XX
RESULT 7
ID ABK11106
XX
AC ABK11106;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding human NOV5a protein, homologue of ankyrin-repeat proteins.
XX
KW Human; NOVX-associated disorder; developmental disorder; blood disorder;
KW endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
KW respiratory disorder; inflammatory disorder; reproductive disorder;
KW neurodegenerative disorder; autoimmune disorder; infectious disease;
KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
KW metabolic pathway modulation; NOV5a; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 6..1061
FT /tag= a
FT /product= "NOV5a"
XX
PN WO200206329-A2.
XX
PD 24-JAN-2002.
XX
PF 18-JUL-2001; 2001WO-US22709.
XX
PR 18-JUL-2000; 2000US-218870P.
PR 18-JUL-2000; 2000US-218875P.
PR 18-JUL-2000; 2000US-218901P.
PR 24-JUL-2000; 2000US-220273P.
PR 26-JUL-2000; 2000US-220912P.
PR 27-JUL-2000; 2000US-221233P.
PR 28-JUL-2000; 2000US-221650P.
XX
PA (CURA-) CUPAGEN CORP.
XX
PI Rastelli L, Shinkets RA, Zerhusen B, Malyankar UM, Padigar M;
XX WPI; 2002-179781/23.
XX

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 8214 BP; 2876 A; 1791 C; 1749 G; 1798 T; 0 other;

Alignment Scores:

Pred. No.: 8,76e-58 Length: 8214
 Score: 672.00 Matches: 155
 Percent Similarity: 57.96% Conservative: 27
 Best Local Similarity: 49.36% Mismatches: 50
 Query Match: 28.68% Indels: 82
 DB: 23 Gaps: 10

US-09-941-831A-20 (1-449) x AAS77257 (1-8214)

QY 1 MetArgGluSerGlyTrpIysLeuIleAspProIleSerAspPheGlyArgMetGlyIle 20
 DB 1879 ATGGGGAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGCGGTATGGGAATA 1820
 QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTrpGluIleCysSerThrTyr 40
 DB 1819 CCCACAGAACTGGACCAATACAGATGCCACAGAACTATGAGATATGCAGCACTTAC 1760
 QY 41 ProProGluIleValValProIysSerValThrLeuGlyThrValValGlySerSerIlys 60
 DB 1759 CCTCTCGAAATAGTGGTTCTTAATCTGTTACCTTGGGAACGGTGGTGGAGTTCAAAG 1700
 QY 61 PheArgSerIysGluArgValProValLeuSerTyrLeuTyrIlysGluAsnAsnAlaA 80
 DB 1699 TTCAGAAAGTAAAGAACGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1640
 QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
 DB 1639 ATTTGGCGGTATGACCGAGCTCTCTCTGATTTTACACTCGCTGTGTATGATGATGAGTC 1580
 QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
 DB 1579 TTGTTGGAGGCCATTAGCCAAACCAACCCAGGAGCCAGTTTATGTTATGTTGTAGACAC 1520
 QY 121 ArgProIysIle----- 124
 DB 1519 AGACCAAGTTGAATGCCATGCCAACCGAGCAGCTGGGAAGGGGTATGAAATGAAGAC 1460
 QY 125 -----TrpHisPheLeu-----ValLeuIleMetArgIle 134
 DB 1459 AACTATGCCAACATTCGTTTCAGATTCATGGGCATTTGAGAACATCCATGATTCGGAGC 1400
 QY 135 ValLeuGlnLeuAlaIysMetAsnLeuMetAspIleThrIysIlePheSerLeuGln 154
 DB 1399 AGTCTGCAG-----AAACTCTTGGAAAGTTGTGCAATTGAAACT 1361
 QY 155 ProAspIysGluGluAspThrAspThrGluGluIysGlnAlaLeuAsn-----Gln 172
 DB 1360 CCAACAAATGAGTGAATTTCTTACGCCCTGGAGAGCTCAGGGTGGTTAAGACACATTA 1301
 QY 173 AlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrIlys 192

DB 1300 GCTATTATGATGCTGGAATTTTC---ATTACAAAGGCAGTGAAGGTAGAAAAGGCCAGT 1244
 QY 193 ArgPheIleAsnSerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaA 212
 DB 1243 GTCTTAGTCCCATTTGTTCTGATGATGGGACCCGACACACAAAGTCTGCTCAGTGGCTAGC 1184
 QY 212 ----- 212
 DB 1183 ATCCTCTAGATCCATTTTATAGGACATTCAAAGGACTCATGAAAGGTGGTATCGTTGT 1124
 QY 213 -----SerTyr-----GlyHisLeuSerCysLeuGlnValLeuLeu 224
 DB 1123 GAGCCAGGTTGTAATTATATCATCAGAAAGTGTGCCACCTC----- 1085
 QY 225 AlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAla 244
 DB 1084 -----GATGGGACTCTAAAGAAGTG-----TCCCTATCTTCACCCAG 1046
 QY 245 ValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
 DB 1045 -----TTCTAGACTGTATCTGGCAATTAATGGAA 1016
 RESULT 9
 ABN60012
 ID ABN60012 standard; cDNA; 1345 BP.
 XX
 AC ABN60012;
 XX
 DT 28-JUN-2002 (first entry)
 DE
 DE Novel human coding sequence SEQ ID NO: 423.
 XX
 KW Human; antianemic; vulnary; antiinflammatory; immunomodulator;
 KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 KW expressed sequence tag; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200222660-A2.
 XX
 PD 21-MAR-2002.
 XX
 PF 10-SEP-2001; 2001WO-US26015.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YR, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 DR WPI; 2002-292408/33.
 DR P-PSDB; ABB97599.
 XX
 PT An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis -
 XX
 PS Claim 1, SEQ ID NO 423; 509pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a coding sequence of the
 CC invention.


```
QY 203 ValProGlyThrProLeuArgLeuAlaAAserTyrGlyHisLeuSerCysLeuGlnVal 222
DB 229 CTCCTCTGCACACCGTTGGATTGGCGCACTGCAGGCATGGAGCTGTGTGGACTTC 288
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
DB 289 CTCATCCGGAAGGGCGCGAGTGGATCTGTGGACGTAAAGGACAGACGCGCCCTGTAT 348
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
DB 349 GTGGCTGTGGTGAACGGGACCTAGAGAGTACCAGATCTCTCGAAGCTGGCGGGAC 408
QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
DB 409 CCCACGGAGCGCGACCATCGCAGCACCCCTGTCTACCGCCTCTCGCTGGCGCG 468
QY 283 ValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
DB 469 GCAGACATCTGAAGCCCTCATCAGGTACGGGCTGTGTTGACGTCAACACACCTTC 528
QY 303 -----ProValTTPAlaSerAsnIleAlaSer-----CysSerGly 314
DB 529 ACTCTGTATGTCAGCTCTGATTCTCCGGCGGCTCACCTCTCTGTGTGTCTGC----- 582
QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
DB 583 CCTTGTATCATCGCAGCCTACCAACCTCCAGTGTTCGGGTGTCTCTCTGGCT 642
QY 335 GlyAlaAspProAspTyrAsnCys-----ThrAspGlnGlyLeuAlaArg 350
DB 643 GCGCGAACCTGACTCAACTGCAATGCTGTCTCAACACACAGGGATTCACAGG--- 699
QY 351 ValProArgProArgThrLeuLeuGluIleCysLeuHisAsnCysGluProGluTyr 370
DB 700 ---GGTCCCTCGGGCGCTCATGGATGCTGTCTCGCCACCGCTGTGAGCAGCCTTC 756
QY 371 IleGlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeu-----ProSerLeuSer 387
DB 757 GTGAGCTCTGTGTAGATTGGAGCAACCTGATCTAGTGAAGTGGGAATGCTGTGGC 816
QY 388 LeuAspLeuThrSerGln-----AspAspLysGlyIleAlaLeuLeuGlnAla 404
DB 817 CCAGAGTCGAGAGGAGAGAAAGTGGACCTCGAGGCTTCGAGCTCTTAAAGAGGCC 876
QY 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgArgAlaLeuCys 424
DB 877 AGAAGTGTTCACAGAACCTTGCTGTCTGTGCGGTGTGCTGTGAGAGAGCTCTTTGGC 936
QY 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyr 444
DB 937 AAA---CACCGGCTTCATCTGATTCCTCTGCTGCTCTGCTGCTGCTGCTGCTGCT 993
QY 445 LeuLysHisGln 448
DB 994 CTACTCCATGAG 1005
RESULT 12
ABV29074
ID ABV29074 standard; cDNA; 1433 BP.
XX
AC ABV29074;
XX
AC ABV29074;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 29065.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
```

```
XX
PD 23-AUG-2001.
XX
PP 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
DR
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 6164-6165; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 1433 BP; 280 A; 377 C; 410 G; 336 T; 30 other;
```

```
Alignment Scores:
Pred. No.: 3,96e-41 Length: 1433
Score: 499.00 Matches: 119
Percent Similarity: 57.89% Conservative: 57
Best Local Similarity: 39.14% Mismatches: 102
Query Match: 21.30% Indels: 26
DB: 23 Gaps: 9
```

```
US-09-941-831A-20 (1-449) x ABV29074 (1-1433)
QY 166 GluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
DB 109 GAGGACACGAGGCTCCATGATGAGCTTACGTGCGGGACCTCCAGACCTCAGAGCCTA 168
QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrp-----Gly 202
DB 169 TTGCAAGAGGAGGAGCTACCGGAGCCGCATCAACGAGAGTGTCTGTGTGTGTGTGTGTG 228
QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
DB 229 CTCCTCTGCACACCGTTGGGAATTCGGGCATCGAGGCATGGGAGCTGTGTGGACTTC 288
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
DB 289 CTCATCCGGAAGGGCGCGAGTGGATCTGTGGACGTAAAGGACAGACGCGCCCTGTAT 348
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
DB 349 GTGGCTGTGGTGAACGGGACCTAGAGAGTACCAGATCTCTCGAAGCTGGCGGGAC 408
QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
```

```
Db 409 CCCACGGAGCGCGCACCATCGCAGCACCCCTGTCTACCCAGCCTCTCGCGTGGCGCG 468
Qy 283 ValAlaIleLeuGlnLeuLeuAseHisGlyAlaGluAlaAseValLysLeu 302
Db 469 GCAGACATCTGAAGCGCCCTCATCAGGTACGGGCGTATGTTGACGTCACACACCTG 528
Qy 303 -----ProValTrpAlaSerAsnIleAlaSer-----CysSerGly 314
Db 529 ACTCTGTATGTCAGCCTCGATCTCCCGCGGCTCACCTCTCTGTGGTCTGC----- 582
Qy 315 ProLeuTrpLeuAlaValTrpGlyHisLeuAseCysPheArgLeuLeuLeuHis 334
Db 583 CCCTTGATCATCAGCGCAGCTACCAACCTCAGTGTCTCCGGCTCTCTCTCTCTCT 642
Qy 335 GlyAlaAseProAseTrpAseCys-----ThrAseGlnGlyLeuLeuAlaAse 350
Db 643 GCGCGAACCCTGACTTCACTCAATGTCGTCTGTCACACACAGGATTTTACAGG--- 699
Qy 351 ValProArgProArgThrLeuLeuGluLeuHisAseCysGluProGluTrp 370
Db 700 ---GGCTCCCTGGGCGCTCATGGATGCTGTCTGCGCCACGGCTGTGAGCGACCTTC 756
Qy 371 IleGlnLeuLeuLeuAsePheGlyAlaAseIleTrpLeu-----ProSerLeuSer 387
Db 757 GTGAGCCTGCTGTAGAAATTTGAGGCAACCTGAATCTAGTGAAGTGGGAATCGCTGGG 816
Qy 388 LeuAseLeuThrSerGln-----AseAsePlyGlyIleAlaLeuLeuGlnAla 404
Db 817 CCAGATCGAGGAGAGAGAAAGTGGACCTTGAGGCGCTTGAGGCTTTTAAAGAGGCC 876
Qy 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
Db 877 AGAAGTGTCCAGAACCTTGTCTGTCTGCGGTGTGCTGTGAGAGAGCTCTTGGC 936
Qy 425 GlnAlaGlyGlnProGlnAlaIleAseGlnLeuAsePileProProMetLeuLeuSerTrp 444
Db 937 AAA---CACCGGCTTCATCTGATTCCTCTGCTGCTCTGCTGCTGCTGCTGCTGCT 993
Qy 445 LeuLysHisGln 448
Db 994 CTACTCCATGAG 1005

RESULT 13
AAD49617
ID AAD49617 standard; cDNA; 1509 BP.
XX AC AAD49617;
XX DT 24-MAR-2003 (first entry)
XX DE Human cytoskeleton-associated protein, CSAP-28 cDNA.
XX KW Human; cytoskeleton-associated protein; CSAP-28; atherosclerosis;
XX KW cancer; gene therapy; gene; ss.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT 12..1025
XX FT /*tag= a
XX FT /product= "Human CSAP-28"
XX PN WO200279404-A2.
XX PD 10-OCT-2002.
XX PF 25-MAR-2002; 2002WO-US09288.
XX PR 29-MAR-2001; 2001US-280508P.
XX PR 03-APR-2001; 2001US-281323P.
XX PR 13-APR-2001; 2001US-283769P.
XX PR 04-MAY-2001; 2001US-288609P.
```

```
PR 10-MAY-2001; 2001US-290518P.
PR 18-MAY-2001; 2001US-291870P.
PR 29-MAY-2001; 2001US-294451P.
XX (INCY-) INCYTE GENOMICS INC.
XX Hafalia AJA, Tang TY, Yue H, Khan FA, Ison CH, Baughn MR;
PI Warren BA, Duggan BM, Thangavelu K, Honchell CD, Azimzai Y;
PI Elliott VS, Burford N, Ding L, Yue H, Becha S, Emerling BM;
PI Richardson TW, Lee SY, Bandman O, Lal PG, Lee S, Gietzen KJ;
PI Wallia NK, Griffin JA, Lee EA, Swarnakar A, Ring HZ, Jones KA;
XX WPI; 2003-092894/08.
DR P-PSDB; AAE32130.
XX New human cytoskeleton-associated proteins, useful for preparing a
PT composition for diagnosing or treating a disease or condition
PT associated with decreased expression or overexpression of functional
PT CSAP e.g., cancer
XX Claim 5; Page 232-233; 233pp; English.
XX The invention relates to new human cytoskeleton-associated protein
CC (CSAP) and its polynucleotide. The polypeptide is useful for preparing
CC a composition for diagnosing or treating a disease or condition
CC associated with decreased expression or overexpression of functional
CC CSAP e.g. atherosclerosis or cancer. The present sequence is human
CC CSAP-28 cDNA. The invention is useful in gene therapy.
XX Sequence 1509 BP; 340 A; 396 C; 449 G; 324 T; 0 other;
SQ
```

```
Alignment Scores:
Pred. No.: 4,27e-41 Length: 1509
Score: 499.00 Matches: 119
Percent Similarity: 57.89% Conservative: 57
Best Local Similarity: 39.14% Mismatches: 102
Query Match: 21.30% Indels: 26
DB: 25 Gaps: 9
```

US-09-941-831A-20 (1-449) x AAD49617 (1-1509)

```
Qy 166 GluLysGlnAlaLeuAseGlnAlaValTrpAseAseAseSerTrpThrLeuAseGlnLeu 185
Db 126 GAGGACAGAGGCTCCATGATCAGTTCAGTTCGGGACCTCCAGACCTCTAGGAGCCTA 185
Qy 186 LeuArgGlnLysArgPheIleAseAseAseAseArgSerGlyTrp-----Gly 202
Db 186 TTGCAAGAGGAGAGTACCGGAGCGCCATCAACAGAGAGTCTGTCTGTGTGTGTGTGTGT 245
Qy 203 ValProGlyThrProLeuArgLeuAlaAseTrpGlyHisLeuSerCysLeuGlnVal 222
Db 246 CTCCTCTGCACACCTTGGGAATCGCGCCACTCGAGCCATCGGAGCTGTGTGAGCTTC 305
Qy 223 LeuLeuAlaHisGlyAlaAseValAseSerLeuAseValLysAlaGlnThrProLeuPhe 242
Db 306 CTCATCCGGAAGGGCGCGAGGTGATCTGTGTGGACGTAAAGGACAGACGCGCTGTAT 365
Qy 243 ThrAlaValSerHisGlyHisLeuAseCysValArgValLeuLeuGlnAlaGlyAse 262
Db 366 GTGGCTGTGTGATGACGCGGACCTAGAGAGTATCCAGATCTCTCGAAGCTCGCGCGAC 425
Qy 263 ProGlyGlySerIleTrpAseAseCysSerProValLeuThrAlaAseAseGlyAla 282
Db 426 CCCAAGGAGCGGCGACCATCGAGCACCTCTGTCTACACGCTCTCTCGCTGGCGCG 485
Qy 283 ValAlaIleLeuGlnLeuLeuAseHisGlyAlaGluAlaAseValLysAlaLysLeu 302
Db 486 GCAGACATCTCTGAAGCGCTCATCAGGTACGGGCTGTGTGTGACGTCAACACACCTG 545
Qy 303 -----ProValTrpAlaSerAsnIleAlaSer-----CysSerGly 314
Db 546 ACTCTGTATGTCAGCCTCGATTTCTCCGCGGCGCTCACCTCTCTCTGTGTGTGTGTGTGT 599
```

QY 315 ProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 659
 QY 335 GlyAlaAspProAspTyrAnCys-----ThrAspGlnClyLeuAlaArg 350
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 716
 QY 351 ValProArgProArgThrLeuLeuGluCysLeuHisHisAsnCysGluProGluTyr 370
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 773
 QY 371 IleGlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeu-----ProSerLeuSer 387
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 833
 QY 388 LeuAspLeuThrSerGln-----AspAspLeuGlyLeuAlaLeuLeuGlnAla 404
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 893
 QY 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 953
 QY 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeuIleSerTyr 444
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 1010
 QY 445 LeuLeuHisGln 448
 Db CTTTGTATCATCAGCGCCTACCAACCTCCAGTGTTCGGGTGCTCTCTCTGGCT 1011

RESULT 14

AAS33289
 ID: AAS33289 standard; cDNA; 1540 BP.

AC AAS33289;
 DT 04-DEC-2001 (first entry)
 XX DNA encoding human secreted protein, Seq ID No 248.

XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
 KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
 KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
 KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
 KW Gaucher's disease; neurological disease; cerebrovascular disorder;
 KW thrombosis; wound healing; ss.

XX Homo sapiens.

XX WO200155326-A2.

XX 17-JAN-2001; 2001WO-US01347.

XX 31-JAN-2000; 2000US-0179065.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-451931/48.

XX P-PSDB; AAU20580.

XX New nucleic acids and polypeptides, useful for diagnosing, preventing
 PT or treating medical conditions -
 PT
 XX
 PS Claim 1; SEQ ID No 248; 753pp; English.
 XX
 XX The invention relates to novel isolated nucleic acid molecules (1)

CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain organs before transplantation, and support cell culture
 CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
 CC coding sequences, PCR primers, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1540 BP; 343 A; 395 C; 454 G; 345 T; 3 other;

Alignment Scores:

Pred. No.: 4,4e-41 Length: 1540
 Score: 499.00 Matches: 119
 Percent Similarity: 57.89% Conservative: 57
 Best Local Similarity: 39.14% Mismatches: 102
 Query Match: 21.30% Indels: 26
 DB: 22 Gaps: 9

US-09-941-831A-20 (1-449) x AAS33289 (1-1540)

QY 166 GluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
 Db 135 GAGGACACGAGGCTCCATGATCGACCTTACGTCGGGACCTCCAGACCTCCAGAGCCCTA 194
 QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrp-----Gly 202
 Db 195 TTGCAAGAGGAGAGCTACCGGAGCGCATCAACAGAGAGTCTGTCTGTGTGGCTGG 254
 QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerGlyHisLeuSerCysLeuGlnVal 222
 Db 255 CTCCCTCGCACACCGTTGGGAATCGGGCCACTCGAGGCCATGGGAGCTGTGTGGACTTC 314
 QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
 Db 315 CTCATCCGGAAGGGGGCGAGGTGGATCTGTGGACGTAAGGACAGACGCGCCCTGTAT 374
 QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
 Db 375 GTGGCTGTGTGAACGGGCACCTAGAGTAGTACCCAGATCTTCTCGAAGCTGGCGCGGAC 434
 QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
 Db 435 CCCAAGCGAAGCGCGCACCATCGAGCCCTGTCTACCCAGCTCTCGGTGGGCGCG 494
 QY 283 ValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
 Db 495 GCAGACATCTCTGAAGGCCCTCATAGGTACGGGGCTGTGACGTCAACACCACTG 554
 QY 303 -----ProValTrpAlaSerAsnIleAlaSer-----CysSerGly 314


```

QY 377 PheGlyAlaAsnIleTyrLeu-----ProSerLeuSerLeuAspLeuThrSerGln 393
Db 821 TTGGAGCCCAACCTGAACCTGGTGAAGTGGGATCCCTGGGCCAGAGGCAAGAGCCAGA 880
QY 394 -----AspAspLysGlyIleAlaLeuLeuLeuGlnAlaArgAlaThrProArgSer 410
Db 881 AGAAGATGGATCCTGAGGCCCTTGCAGGTCTTTAAAGAGGCCCAAGATATCCAGGACC 940
QY 411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 941 TTGCTGAGTTGTGCCGGGTGGCTGTGAGAGAGACTCTTGGCAATAC---CGACTGCAT 997
QY 431 AlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyrLeuLysHisGln 448
Db 998 CTGGTCCCTCGCTGCCGCTGCCAGACCCCAAGAGCTTTTGTGCTTTATGAG 1051

```

Search completed: December 1, 2003, 08:00:21
Job time : 377 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:53:58 ; Search time 2294 Seconds
(without alignments)
644.863 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGKLIPDIFGRMG.....QAINQLDIPPLMLSYLKHQL 449

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Fgapop 10.0 , Fgapext 0.5
Delop 6.0 , Delext 7.0

Searched: 219069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xl
-MODEL=frame_p2n.model -DEV=xl

-Q=/cgn2_1/USPTO.spool_p/US0941831/runat_01122003_073528_7884/app_query.fasta_1.647
-DB=Published Applications NA -QMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCL=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0941831_@CNG_1_1_85 @runat_01122003_073528_7884
-NCPUS=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PTC_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PTC_NEW_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2343	100.0	1878	10	US-09-941-831-6 Sequence 6, Appli

2	2343	100.0	1878	12	US-10-207-175-13
3	1690	72.1	1222	10	US-09-908-193-13
4	1629	69.5	1069	10	US-09-908-193-11
5	606.5	25.9	522	10	US-09-833-381-62
6	480	20.5	2019	10	US-09-908-805B-24
7	265	11.3	1118	12	US-10-314-321A-60
8	265	11.3	1251	12	US-10-314-321A-64
9	264	11.3	3482	11	US-09-919-039-133
10	261.5	11.2	1206	14	US-10-037-270-34
11	246	10.5	990	9	US-09-884-870-3
12	246	10.5	1538	9	US-09-884-870-1
13	244.5	10.4	401	9	US-09-923-217-28
14	244.5	10.4	401	10	US-09-833-263-28
15	244.5	10.4	401	13	US-10-025-380-28
16	233.5	10.0	389	10	US-09-867-701-5266
17	233.5	10.0	419	10	US-09-908-805B-26
18	230	9.8	549	12	US-10-314-321A-58
19	230	9.8	616	12	US-10-314-321A-59
20	223	9.5	5482	12	US-10-199-937-138
21	217.5	9.3	4275	11	US-09-972-115A-5
22	217.5	9.3	4406	12	US-10-199-937-106
23	217.5	9.3	4992	12	US-10-199-937-134
24	217.5	9.3	5002	12	US-10-199-937-132
25	217.5	9.3	6018	12	US-09-849-602-11
26	216	9.2	6096	12	US-10-133-013-169
27	215	9.2	1568	12	US-10-199-937-98
28	215	9.2	3400	9	US-09-509-196A-1
29	215	9.2	3508	12	US-10-199-937-1
30	215	9.2	4127	12	US-10-199-937-100
31	215	9.2	5778	14	US-10-163-587A-14
32	214	9.1	2971	12	US-10-199-937-92
33	214	9.1	3353	12	US-10-199-937-88
34	214	9.1	3799	12	US-10-199-937-90
35	213.5	9.1	4491	9	US-09-841-835-7
36	209	8.9	4657	9	US-09-841-835-9
37	208.5	8.9	3984	12	US-10-199-937-3
38	208.5	8.9	4134	9	US-09-841-835-1
39	208.5	8.9	4134	11	US-09-972-115A-7
40	207	8.8	1712	14	US-10-037-270-1101
41	205	8.7	5175	10	US-09-964-899-42
42	205	8.7	14770	12	US-09-873-367C-230
43	203	8.7	6207	12	US-10-205-194-118
44	200	8.5	1066	9	US-09-835-788A-8
45	200	8.5	1066	12	US-10-175-042-8

ALIGNMENTS

RESULT 1
US-09-941-831-6
; Sequence 6, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: PTO4991
; FILE REFERENCE: Serine/Threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941,831
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-941-831-6

Alignment Scores: 1.54e-298 Length: 1878
Pred. No.: 2343.00 Matches: 449
Score:

QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
DB 547 TTCAGAGTAAGAAGCTGCTCCTGTCTCTCTACCTCTCAAGAGACAACTGCTGCC 606
QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
DB 607 ATTTGCGGCTGAGCAGCCTCTCTCTGATTTTACACTCGCTGTGTAGATGATGAGCTC 666
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
DB 667 TTGTTGGAGGCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA 726
QY 121 ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
DB 727 AGACCAAGATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCATTATGCCAAG 786
QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
DB 787 ATGAACCTCATGGACATCAACCAAGATCTTCTCCCTCTGAGCCGACCAAGGAGGAG 846
QY 161 AspThrAspThrGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
DB 847 GACACTGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACACTCTTAT 906
QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
DB 907 ACTTTGGACCACTTTTGGCCAGGAGCGTTTACAACGTTTCTATCAACGACGAGTGGC 966
QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeu 220
DB 967 TGGGGTGTCTCTGGGACACCTTGGCTTGGCTGCTTCTTATGGCCTTGTAGCTGTTTG 1026
QY 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240
DB 1027 CAAGTCCTCTTAGCCCATGGCTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCCA 1086
QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
DB 1087 CTTTTCACCTGCTGTCAGTCATGCCCATCTGGACTGTGTACGTGCTTTTGAAGCTGT 1146
QY 261 AlaSerProGlyCysSerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
DB 1147 GCCTCTCTCTGGTGTAGCATCTACAACACTGTCTCCGCTGCTCACAGCTGCCGTGAT 1206
QY 281 GlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
DB 1207 GGTGCTGTTGCTATCTCGAGGAGCTCTTAGACCATGGTGTGACAGGCCCAACGCTCAA 1266
QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320
DB 1267 AAACCTACAGTCTGGGCATCAAACTAGCTTCTGCTGCTGCTCACAGCTGCCGTGAT 1326
QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
DB 1327 GTCTAGGGGACCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGACACCTGACTAC 1386
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeu 360
DB 1387 AACTGACCTGACAGGCGCTATTGGCTGCTGCTCCCAAGACCCCGCACCTCTCTGAAATC 1446
QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
DB 1447 TGCCTCCATCAATATGTGAGCCAGATATATCCAGCTGTATGATTTTGGTGTCTAT 1506
QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
DB 1507 ATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGCATTGCAATG 1566
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
DB 1567 CTGCTACAGGCGGAGCCACTCCAGGTCACCTTCTATCATCAGGTCGTTTGTAGTCGTCGC 1626

QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMet 440
DB 1627 AGAGCTTTGTCCAGGCTGGCCAGCCACCAAGCCATCAACCAAGCTGATATTCTCCCATG 1686
QY 441 LeuIleSerTyrLeuLysHisGlnLeu 449
DB 1687 TTGATTAGCTACCTAAAACACCAACTG 1713
RESULT 3
US-09-908-193-13
; Sequence 13, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1222
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
; OTHER INFORMATION: with homology to ankyrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-13
Alignment Scores:
Pred. No.: 1,55e-212 Length: 1222
Score: 1690.00 Matches: 326
Percent Similarity: 99.70% Conservative: 1
Best Local Similarity: 99.99% Mismatches: 1
Query Match: 72.13% Indels: 0
DB: 10 Gaps: 0
US-09-941-831a-20 (1-449) x US-09-908-193-13 (1-1222)
QY 122 ProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMet 141
DB 102 CCACAGATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCAATTAGCCAAAGATG 161
QY 142 AsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAsp 161
DB 162 AACCTCATGGACATCAACCAAGATCTTCTCCCTCTGAGCCGACCAAGGAGGAGGAC 221
QY 162 ThrAspThrGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThr 181
DB 222 ACTGACACAGAGAGAGAGGCTTCTCAATCAAGCAGTGTATGACACGACTCTCTATAT 281
QY 182 LeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyr 201
DB 282 TTGGACCAAGCTTTTGGCCAGGAGGCTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGG 341

QY 202 GlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGln 221
Db 342 GGTGTTCTCGGACACCCCTGGCTGGCTCTTATGGCCACTTGAGCTGTTGGCAA 401
QY 222 ValLeuAlaHisGlyAlaAspValAspSerLeuAspValysAlaGlnThrProLeu 241
Db 402 GTCTCTTAGCCCATGGTCTGATGTGACAGCTTGATGTCAAGGCACAGCCACTT 461
QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261
Db 462 TTCACTGCTGTGAGTCAATGCCATCTGGACTGTGTACGTGTCTTTGGAAGCTGGTGC 521
QY 262 SerProGlyGlySerIleTyrHisLeuSerProValLeuThrAlaAlaArgAspGly 281
Db 522 TCTCTGGGTGGTAGCATCTACAACTGTCTCCGCTGTCTCACAGCTGCCGTGATGT 581
QY 282 AlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValysAlaLys 301
Db 582 GCTGTGTCTATCTCGAGGAGCTCTAGACCATGGTGCAGAGCCCAACGTCAAAGCTAA 641
QY 302 LeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaVal 321
Db 642 CTACCACTCTGGGCATCAAACTAGCTTCATGTTCTGGCCCTCTATTTGGCCGAGTC 701
QY 322 TyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn 341
Db 702 TAGGGGCCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGCAGACCTGACTACAAC 761
QY 342 CysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCys 361
Db 762 TGCACTGACAGGGCTTATTTGGCTGTGTCCTCCAGACCCCGCACCTCTTGAATCTGC 821
QY 362 LeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381
Db 822 CTCCATCATATTTGAGCCAGAGTATATCCAGCTGTAAATGATTTTGGTGTCAATATC 881
QY 382 TyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspGlyGlyIleAlaLeuLeu 401
Db 882 TACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGGCATTCATTTGCTG 941
QY 402 LeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgArg 421
Db 942 CTACAGCCCGAGCCACTCCAGGTCACTTCTATACAGCTCGTTCGTTCGCGACA 1001
QY 422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeu 441
Db 1002 GCCTTGTGCAGCTGCCAGCCACAAAGCCATCAACAGCTGGATATCTCCCATGTTG 1061
QY 442 IleSerTyrLeuLysHisGlnLeu 449
Db 1062 ATTAGCTACCTAAACACCAACTG 1085

RESULT 4

US-09-908-193-11
; Sequence 11, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1069
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
; OTHER INFORMATION: with homology to ankryrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-11

Alignment Scores:

Pred. No.: 1.43e-204 Length: 1069
Score: 1629.00 Matches: 318
Percent Similarity: 97.55% Conservative: 0
Best Local Similarity: 97.55% Mismatches: 0
Query Match: 69.53% Indels: 8
DB: 10 Gaps: 1

US-09-941-831A-20 (1-449) x US-09-908-193-11 (1-1069)

QY 132 MetArgIleValLeuGlnLeuAlaLysMetAsnLeuMetAspIleThrIysIlePheSer 151
Db 81 ATGAGAAATAGTTCCTCCAATTAGCCAAAGATGAACCTCATGGACATCACCAGATCTCTCC 140
QY 152 LeuLeuGlnProAspLysGluGluAspThrAspThrGluGluLysGlnAlaLeuAsn 171
Db 141 CTCCTGACGCCGCAAGGAGGAGGAGGACACTGACACAGAGGAGAGCAGCTCTCAAT 200
QY 172 GlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyr 191
Db 201 CAAGCAGTGTATGACACAGACTCTCTATATTTGGACCACTTTTGGCCAGAGCGTTAC 260
QY 192 LysArgPheIleAsnSerArgSerGlyTyrGlyValProGlyThrProLeuArgLeuAla 211
Db 261 AAACGTTTTCATCAACAGCAGAGAGTGGCTGGGGTGTCTTGGACACACCTTGGCTGGCT 320
QY 212 AlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 231
Db 321 GCTTCTTATGGCCACTTGAGCTGTTTGAAGTCTCTTTAGCCCATGGTGTCTGATGTTGAC 380
QY 232 SerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp 251
Db 381 AGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGTCTGTCTGATGCGCATCTGGAC 440
QY 252 CysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrHisLeuAsnCys 271
Db 441 TGTGTACGTGTGCTTTTGGAGCTGGTGCCTCTCTCTGGTGGTAGCATCTACAACTGT 500
QY 272 SerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAsp 291
Db 501 TCTCCGCTGCTCACAGCTGCCGCTGATGGTGTGTGTCTATCTCTCAGAGAGCTCTAGAC 560
QY 292 HisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrpAlaSerAsnIleAlaSer 311
Db 561 CATGCTGACAGAGCCCAAGCTCAAAAGCTAACTACCACTGTGGGCATCAAACTAGCTTCA 620
QY 312 CysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeu 331
Db 621 TGTGTGGCCCTCTATTTGGCCGAGTCTACGGGACCTTGGACTGTGTTCGCCCTGCTT 680
QY 332 LeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgVal 351
Db 681 TTGCTCCAGGGGAGAGCCCTGACTACAACCTGCACTGACAGGCGCTATTGGCTGTGTC 740


```
Db 287 CCCTAGGATCGAGCCACTGCGAGCCATGGGAACCTGTGTGGACTTCCTCATACGCAAA 346
Qy 227 GlyAlaAspValAspSerLeuAspValValysAlaGlnThrProLeuPheThrAlaValSer 246
Db 347 GGGCCGAGGTGACCTGGTGGATGCAAGGGCAGACTGCCCTGTATGTGGTGTAGTG 406
Qy 247 HisGlyHisLeuAspCysValArgValLeuGluAlaGlyAlaSerProGlyGlySer 266
Db 407 AACGGGCACCTGGAGACACTGAGATCCTTTTGAAGCTGGTGTGATCCCAACGGCAGC 466
Qy 267 IleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLeu 286
Db 467 CGCACACACCGACACTCCTCGTGTACCATGSCCTTCGTGTGGTAGGACGACATCCTG 526
Qy 287 GlnGluLeuAspHisGlyAlaGluAlaAsnValysAlaLeu 302
Db 527 AAGGCTCTTATCAGGTATGGGCGAGATGTGTGATCATCATCTGAAATTCGTACACC 586
Qy 303 ---ProValTrpAlaSerAsnIleAlaSer-----CysSerGlyProLeuTyrLeu 318
Db 587 CGGCCCCCTTTTTCACGGCGGTAACCTCCTTGGTGGTCTGT-----CCTCTATACATC 640
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 641 AGTGTGCTTACCATACATACCTTCACTTGTGCTTGTGAGGCTGTCTTGTGAGGCTGGGCAATCCT 700
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArg-----ProArgThr 356
Db 701 GACTTCANATGCAATGGCCCTGTACACCCAGGAGTTCTACAGGGATCCCTCGGGTGT 760
Qy 357 LeuLeuGluLeuCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAsp 376
Db 761 GTCATGATGCTGCTCTCGCCCATGGCTGTGTAAGCAGCCCTTCGTGAGTCTGTTGGTAGAG 820
Qy 377 PheGlyAlaAsnIleTyrLeu-----ProSerIleSerLeuAspLeuThrSerGln 393
Db 821 TTTGGAGCCCAACCTGAACCTGGTGAAGTGGGAATCCCTGGGCCAGAGGCAAGAGCAGA 880
Qy 394 -----AspAspLysGlyIleAlaLeuLeuGlnAlaLeuAlaArgAlaThrProArgSer 410
Db 881 AGAAAGATGATCTGAGGCTTGCAGGTCTTTAAAGAGCCAGAGATATCCAGGACC 940
Qy 411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 941 TTGCTGAGTTTGTGCGGGTGGCTGTGAGAAGACTCTTGGCAATAC---CGACTGCAT 997
Qy 431 AlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyrLeuLysHisGln 448
Db 998 CTGGTTTCTCGCTGCTGCGCTGCCAGACCCCAATAAAGAAGATTTTGTCTTTATGAG 1051
```

RESULT 7

```
US-10-314-321A-60
; Sequence 60, Application US/10314321A
; Publication No. US20030190648A1
; GENERAL INFORMATION:
; APPLICANT: Hitachi, Ltd.
; TITLE OF INVENTION: Gene Predicting Method
; FILE REFERENCE: 31010118US1
; CURRENT APPLICATION NUMBER: US/10/314,321A
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: JP 2002-103333
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1118
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: partial sequence of ALJ65356
US-10-314-321A-60
```

Alignment Scores:

```
Pred. No.: 3,13E-24 Length: 1118
Score: 285.00 Matches: 85
Percent Similarity: 47.15% Conservative: 39
Best Local Similarity: 32.32% Mismatches: 96
Query Match: 11.31% Indels: 44
DB: 12 Gaps: 10
```

US-09-941-831A-20 (1-449) x US-10-314-321A-60 (1-1118)

```
Qy 197 SerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHis 216
Db 32 ACAGGTACCCCTGG---ACTCCATC-ACGCCCTGACGACGCCAGCTGTGAGGCCAG 87
Qy 217 LeuSerCysLeuGlnValLeuAlaHisGlyAlaAspValAspSerLeuAspValys 236
Db 88 GCGCGTGTGTGAGCTGTCTGCGGCTGGGCGCCAGTGTGATGCTCGCAACATCGAC 147
Qy 237 AlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeu 256
Db 148 GGCAGCACCCCTCTCGCATGCTCGGCTCGGGCAGCATCGAGTGTGTGAAGCTCTTG 207
Qy 257 LeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThr 276
Db 208 CTGTCTACCGGGCCCAAGGTCAACCTCCCTCTGTAC---ACAGCGTCCCCCTGCACGAG 264
Qy 277 AlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAla 296
Db 265 GCTGTCATGAGCGGAGTTCGGAATGTGTGAGGCTCTTATTGACGTCGGG---GCC 318
Qy 297 AsnValLysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCys-----SerGly 314
Db 319 AATCTGGAAGCGCAC-----GATTGCCATTTTGGGACC 351
Qy 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
Db 352 CCTCTGCACTGTGCTGTGCGGAGCATCTGGACTGTGTCAAAGTGTCTCAATGCA 411
Qy 335 GlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
Db 412 GGGGCCAAGCTGAAT-----GCG 429
Qy 355 ArgThrLeuLeuGluLeuCysLeuHisHisAsnCysGlu-----ProGluTyrIle 371
Db 430 GCAAAAGCTTATGAGACTGCCCCCTTCAACACGCGGCCAAGGTCAAGAAATGTTGACCTCATC 489
Qy 372 GlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeu-----Pro 384
Db 490 GAGATGCTTATCGAGTTTGGCGGCAACATCTACGCCCGGACACCCGCGGGAAGACCG 549
Qy 385 SerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuGlnAla 404
Db 550 TCT-----GACTACAGTGGAGCAGCAGCTCCGCCCAAGTGTTCGAGTACTAC 600
Qy 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
Db 601 GAAAAGACACCTCTGACTCTGTACAGCTCTGTGAGGCTGAACTTGAAGGAGCCACTGCG 660
Qy 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyr 444
Db 661 GTCCGAGGGGCTG---GAGAAGATTCCCAAGTTAAACATCCCGCCCGGCTCATTTGATTAC 717
Qy 445 LeuLysHis 447
Db 718 CTCTCCTAC 726
```

RESULT 8

```
US-10-314-321A-64
; Sequence 64, Application US/10314321A
; Publication No. US20030190648A1
; GENERAL INFORMATION:
; APPLICANT: Hitachi, Ltd.
; TITLE OF INVENTION: Gene Predicting Method
```

QY	425	GlnAlaGlyGlnProGlnAlaAlaAsnGlnLeuAspIleProMetLeuLeuSerTyr	444
Db	694	GTCCAGGGCTG---GAGAGATTGCCAAGTTAAACATCCCGCCCGCTCATTTGATTAC	750
QY	445	LeuLysHis	447
Db	751	CTCTCTCTAC	759
RESULT 9			
US-09-919-039-133			
; Sequence 133, Application US/09919039			
; Publication No. US20030108871A1			
; GENERAL INFORMATION:			
; APPLICANT: Kaser, Matthew R.			
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES			
; FILE REFERENCE: PA-0035 US			
; CURRENT APPLICATION NUMBER: US/09/919,039			
; CURRENT FILING DATE: 2002-09-09			
; PRIOR APPLICATION NUMBER: 60/222,113			
; PRIOR FILING DATE: 2000-07-28			
; NUMBER OF SEQ ID NOS: 401			
; SOFTWARE: PERL Program			
; SEQ ID NO 133			
; LENGTH: 3482			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc feature			
; OTHER INFORMATION: Incyte ID No. US20030108871A1 1823159CB1			
US-09-919-039-133			
Alignment Scores:			
Pred. No.: 2,938-23 Length: 3482			
Score: 264.00 Matches: 80			
Percent Similarity: 42.00% Conservative: 46			
Best Local Similarity: 26.67% Mismatches: 106			
Query Match: 11.27% Indels: 68			
DB: 11 Gaps: 12			
US-09-941-831A-20 (1-449) x US-09-919-039-133 (1-3482)			
QY	5	GlyTrrpLysLeuLeuAspProIleSerAspPheGlyArgMetGlyIleProAsnArgAsn	24
Db	600	GGATGGACAGTTTACATCCAGTGGAGAAATACAGGAGGAGGCTTGGCCCAATCACCAT	659
QY	25	TrrpThrIleThrAspAlaAsnArgAsnTyrGluLeuCysSerThrTyrProProGluIle	44
Db	660	TGGAGAAATAACTTTTATTAATAAGTGCTATGAGCTCTGTGACACTTACCCCTGCTCTTTG	719
QY	45	ValValProLysSerValThrLeuGlyThrValValGlySerSerLysPheArgSerLys	64
Db	720	GTGGTTCGTATCTGTCGCTCAGATGATGACCTCCGGAGAGTTGCAACTTTTAGGTCCCGA	779
QY	65	GluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAlaAlaIleCysArgCys	84
Db	780	AATCGAATTCAGTCTGTCTGATGATTCATCCAGAAATAAGACGGTCAATTTGCGTTGC	839
QY	85	SerGlnProLeuSerGlyPheTyr---ThrArgCysValAspAspGluLeuLeuGlu	103
Db	840	AGTCAGCCTCTTGTGCGTATGAGTGGGAAAGAAATAAAGATGATGAGAAATATCTCGAT	899
QY	104	AlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThrArgProLys	123
Db	900	GTATCAGGGAGACTAATAACAATTTCTAACTCACCATTATGATGCAAGACCAGC	959
QY	124	Ile-----	126
Db	960	GTAAATGCAGTGGCCCAACAGGAGGAGGATATGAAAGTGATGATGATGATCATCAT	1019
QY	127	PheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMetAsnLeuMet	144
Db	1020	AACGCGCACTTTTCTTC-----TTAGACATTTAATAATATTCATTTATGTTGGGAA	1070

; CURRENT FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/212,222
 ; PRIOR FILING DATE: 2000-06-16
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 972
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(972)
 ; US-09-884-870-3

Alignment Scores:
 Pred. No.: 8,27e-22 Length: 990
 Score: 246.00 Matches: 82
 Percent Similarity: 45.99% Conservative: 44
 Best Local Similarity: 29.93% Mismatches: 100
 Query Match: 10.50% Indels: 48
 DB: 9 Gaps: 9

US-09-941-831A-20 (1-449) x US-09-884-870-3 (1-990)

QY 206 ThrProLeuArgLeuAlaalaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
 Db 199 TCCCACTTCATGAAGCTCAGCTCAGGGCGCTTACTGGCCCTTAAACTTTAATTGCA 258
 QY 226 HisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaVal 245
 Db 259 CAAGGTGTCAATGTGAACCTTGTGACAAATTAACCGGGTGTCTCTCCACGAGGCATGC 318
 QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
 Db 319 CTGGAGGTCAGCTGCTGCGCCATGTCGCAAGCTTATTGGAAATGGTGCAACGCTCAATGGA 378
 QY 266 SerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLle 285
 Db 379 GTGACAGTTCAGGAGCCACACCCCTTCAATGCTTGTGTCAGCGCAGTGTGATGT 438
 QY 286 LeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrrp 305
 Db 439 GTCAATGTGTCTGTGAGTTCGAGGCAAGGCCAGTTCGAGGTGCACCTG----- 489
 QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeu 325
 Db 490 -----GCCCGCCCATCCATGAGGCAGTGAAGAGAGGTGCACAGA 528
 QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn----- 341
 Db 529 GAGTGCATGGAGATCTCTGTGGCAATATATTGAACATTGACCATGAGGTGCTCAGCTC 588
 QY 342 -----CysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
 Db 589 GGAACCTCCCTATATGTGGCTGCACCTACAG-----AGGTAGACTGTGTG 636
 QY 355 ArgThrLeuLeuLeuLeu-----CysLeu 362
 Db 637 AGAAACTCTTAGAATTAGAGCCAGTGTGACCATCGCCAGTGGCTGGACACCCCACTC 696
 QY 363 His-----HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAla 379
 Db 697 CATGTGTCAGCGAGGAGTCCCATGTGGAGGTCACTCACCCTGTACCACTATGGAGCT 756
 QY 380 AsnIleTyrLeuProSerLeu-----SerLeuAspLeuThrSer---GlnAsp 394
 Db 757 AACCTGAAGCGTAGAATGTCTAGGCAAAAGTGCCTTGTATCTGCGCGCTCCAAAAGC 816
 QY 395 AspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGln 414
 Db 817 AGCGTGGAGCAGGCACTCTTGCTC-----CGTGAAGGCCCACTCTCTCTTTCCCACTC 870
 QY 415 ValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGln 434

Db 871 TGCCGCCTGTGTGTCGGAAGTGTCTCGTGCAGCATGTCTAT---CAAGCCATCCACAG 927
 QY 435 LeuAspIleProProMetLeuIleSerTyrLeuLysHisGln 448
 Db 928 CTACATCTGCCAGAGCCCACTGACAGGATTCCTCTTATACCAA 969

RESULT 12

US-09-884-870-1
 ; Sequence 1, Application US/09884870
 ; Patent No. US20020065223A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glucksmann, Maria A.
 ; APPLICANT: Kadambi, Vivek
 ; TITLE OF INVENTION: 33358, A NOVEL HUMAN ANKYRIN FAMILY MEMBER AND USES THEREOF
 ; FILE REFERENCE: MNI-162CP
 ; CURRENT APPLICATION NUMBER: US/09/884,870
 ; CURRENT FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/212,222
 ; PRIOR FILING DATE: 2000-06-16
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1538
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (75)...(1046)
 ; US-09-884-870-1

Alignment Scores:
 Pred. No.: 1,75e-21 Length: 1538
 Score: 246.00 Matches: 82
 Percent Similarity: 45.99% Conservative: 44
 Best Local Similarity: 29.93% Mismatches: 100
 Query Match: 10.50% Indels: 48
 DB: 9 Gaps: 9

US-09-941-831A-20 (1-449) x US-09-884-870-1 (1-1538)

QY 206 ThrProLeuArgLeuAlaalaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
 Db 273 TCCCACTTCATGAAGCTCAGCTCAGGGCGCTTACTGGCCCTTAAACTTTAATTGCA 332
 QY 226 HisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaVal 245
 Db 333 CAAGGTGTCAATGTGAACCTTGTGACAAATTAACCGGGTGTCTCTCCACGAGGCATGC 392
 QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
 Db 393 CTGGAGGTCAGCTGCTGCGCCATGTCGCAAGCTTATTGGAAATGGTGCAACGCTCAATGGA 452
 QY 266 SerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLle 285
 Db 453 GTGACAGTTCAGGAGCCACACCCCTTCAATGCTTGTGTCAGCGCAGTGTGATGT 512
 QY 286 LeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrrp 305
 Db 513 GTCAATGTGTCTGTGAGTTCGAGGCAAGGCCAGTTCGAGGTGCACCTG----- 563
 QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeu 325
 Db 564 -----GCCCTGCCCATCCATGAGGCAGTGAAGAGAGGTGCACAGA 602
 QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn----- 341
 Db 603 GAGTGCATGGAGATCTCTGCTGCAATAATATTGAACATTGACCATGAGGTGCTCAGCTC 662
 QY 342 -----CysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
 Db 663 GGAACCTCCCTATATGTGCGCTGCACCTACCA-----AGGTAGACTGTGTG 710

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:56 ; Search time 44 Seconds
(without alignments)
981.358 Million cell updates/sec

Title: US-09-941-831A-20

Perfect score: 2343

Sequence: 1 MRESGWKLIDPISDFGRMGI.....QAINQLDIPPLMISYLKHQL 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313.5	13.4	676	2 T22532	hypothetical prote
2	224.5	9.6	1848	2 S37771	ankyrin, erythrocy
3	224.5	9.6	1862	2 I49502	ankyrin - mouse
4	224.5	9.6	1881	1 SJHUK	ankyrin 1, erythro
5	221.5	9.5	1856	2 B35049	ankyrin 1, erythro
6	221.5	9.5	1880	2 A35049	ankyrin 1, erythro
7	218.5	9.3	2039	2 T15347	ankyrin-related un
8	214.5	9.2	1549	2 T13940	ankyrin - fruit fl
9	212.5	9.1	559	2 T37958	probable myotubula
10	212	9.0	1062	2 T14151	inv protein - mous
11	210.5	9.0	688	2 S57131	hypothetical prote
12	208	8.9	1062	2 T30255	inversin - mouse
13	206.5	8.8	1765	2 T42714	ankyrin 3, splice
14	206.5	8.8	1940	2 T42715	ankyrin 3, splice
15	206.5	8.8	1943	2 T42713	ankyrin 3, splice
16	206.5	8.8	1961	2 T42716	hypothetical prote
17	205	8.7	276	2 T12477	ankyrin 3, long sp
18	205	8.7	4377	2 A55575	hypothetical prote
19	203.5	8.7	1435	2 T32930	ankyrin 2, neurona
20	202.5	8.6	3824	2 S37431	hypothetical prote
21	199.5	8.5	557	2 T46507	hypothetical prote
22	196.5	8.4	426	2 AE2149	hypothetical prote
23	196.5	8.4	1031	2 T43458	hypothetical prote
24	191	8.2	368	2 T18184	ankyrin repeat pro
25	187	8.0	410	1 WWVZF7	BamHI-ORF7 protein
26	181	7.7	799	2 T48451	myotubularin-like
27	181	7.7	845	2 T07052	probable potassium
28	180.5	7.7	791	2 T42691	hypothetical prote
29	177.5	7.6	426	2 E84689	probable RING zinc

30	177	7.6	887	2 T03939	potassium channel
31	174.5	7.4	888	2 D84650	probable potassium
32	171.5	7.3	2584	2 T24158	hypothetical prote
33	171.5	7.3	2606	2 T24157	hypothetical prote
34	169.5	7.2	838	2 S23606	potassium channel
35	169.5	7.2	857	2 S62694	potassium channel
36	169	7.2	262	2 T17495	ankyrin repeat pro
37	168.5	7.2	679	2 B45771	2-5A-dependent RNA
38	168	7.2	247	2 D84448	probable ankyrin l
39	167	7.1	323	2 B47169	ankyrin-like repea
40	166	7.1	1398	2 T21884	hypothetical prote
41	165	7.0	397	2 T46445	hypothetical prote
42	164.5	7.0	1066	2 S72479	Nuc-2 protein - Ne
43	164	7.0	807	2 T12177	potassium channel
44	161	6.9	180	2 A75579	ankyrin-related pr
45	160.5	6.9	1411	2 S30355	alpha-latroinsecto

ALIGNMENTS

RESULT 1

T22532

hypothetical protein F53A2.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22532

R:Lloyd, C.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19576

A:Accession: T22532

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-676 <WIL>

A:Cross-references: EMBL:Z81546; PIDN:CAB04456.1; GSPDB:GN00021; CESP:F53A2.8

A:Experimental source: clone F53A2

C:Genetics:

A:Gene: CESP:F53A2.8

A:Map position: 3

A:Introns: 10/3; 78/3; 104/1; 419/3; 449/3; 477/3; 652/3

Query Match 13.4%; Score 313.5; DB 2; Length 676;

Best Local Similarity 35.3%; Pred. No. 9.6e-18;

Matches 59; Conservative 37; Mismatches 48; Indels 23; Gaps 2;

QY 5 GWKLIDPISDFGRMCIPIRNRWTTITDANRYEICSTYPEIIVVPKSVTLGTGVGSSKFRSK 64

DB 125 GWRRLDWNSEMTROGITKRSQWTESINEGYTICDTYPNKLWFPPTAASTSVLLGSCCKFRSR 184

QY 65 ERVPVLSYLYKENNAIICRCQPLSGFYTRCVDDLELLEAISQTPNGSQFMVYVDTTRPKI 124

DB 185 GRLPVLTTFHQOTEALRCACQPLTGFSAFCVDEKLMELVGKANTNSDNLFLVDTRPKR 244

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

QY 125 -----WHLVLMIRIVLQAKNNLMD-ITK 148

DB 245 NAMVNKVGKGFEDERNYSNMRFFFDIENIHVMRASQARLLDAVTK 291

F;267-299/Domain: ankryn repeat homology <AN08>
F;300-332/Domain: ankryn repeat homology <AN09>
F;333-365/Domain: ankryn repeat homology <AN10>
F;366-398/Domain: ankryn repeat homology <AN11>
F;399-431/Domain: ankryn repeat homology <AN12>
F;432-464/Domain: ankryn repeat homology <AN13>
F;465-497/Domain: ankryn repeat homology <AN14>
F;498-530/Domain: ankryn repeat homology <AN15>
F;531-563/Domain: ankryn repeat homology <AN16>
F;564-596/Domain: ankryn repeat homology <AN17>
F;597-629/Domain: ankryn repeat homology <AN18>
F;630-662/Domain: ankryn repeat homology <AN19>
F;663-695/Domain: ankryn repeat homology <AN20>
F;696-728/Domain: ankryn repeat homology <AN21>
F;729-761/Domain: ankryn repeat homology <AN22>
F;762-794/Domain: ankryn repeat homology <AN23>

Query Match 9.6%; Score 224.5; DB 2; Length 1862;
Best Local Similarity 32.2%; Pred. No. 8.7e-10;
Matches 76; Conservative 37; Mismatches 86; Indels 37; Gaps 7;

Qy 169 ALNOAVVNDNSYTLQLLRQRKYRKFINSRGWGPVTPLRLAASYGHLSCLOVLLAHGA 228
Db 172 ALHIAARNDDTRTAALLQNDFNPDVL-SKTGF-----TFLHTAAHYENLNVAQLLNRGA 226

Qy 229 DVDSLDVKAOPTPLFTAVSHGHDLDCVRVLEAGSPGSIYNNCSPVLTTAADGAVAILEO 288
Db 227 SVNFTPQGITPLHIASRRGNVINRVLLLDRCAQTETRTKDELTPLHCAAARNGHVRISEI 286

Qy 289 LLDHGAENVAKUPLVMNASNTASSCGPIYLAAVGYHDLCFRLLLLHGADPDYNYCTD---- 344
Db 287 LLDHGAPIQATK-----NLGSPIHMRAQQGDHLDCCVRLLLQYNAEIDDTILDHLTLP 337

Qy 345 -----QGLLARV-----PRPRTLLEIC-LHHNCEPEXI---OLLIDFGANI 381
Db 338 LHVAACHGHHRVAKVLLDKGAPKNRSALNGFTPLPHIAACKNQHNIRVNELLTKTGASI 393

RESULT 4
SJHUUK
ankyrin l, erythrocyte splice form 1 - human
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
N;Contains: ankyrin 2.2
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 22-Jun-1999
C;Accession: S08275; A33219; FC2220; A35443
R;Lux, S.E.; John, K.M.; Bennett, V.
Nature 344, 36-42, 1990
A;Title: Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure w
A;Reference number: S08275; UID:90158830; PMID:2137557
A;Accession: S08275
A:Molecule type: mRNA
A;Residues: 1-1881 <LU>
A;Cross-references: EMBL:X16609; NID:g28701; PIDN:CAB4610.1; PID:g28702
A;Accession: A33219
A:Molecule type: protein
A;Residues: 2-7,'X','9-17,'X',19-20,'T',22-30:'733-749','A',751-753;828-833,'X',835-855,'X',
X',1367;1383-1427;1601-1630;1686-1698,'D',1700;'1763-1772<LUX>
A>Note: 845-Arg and I392-Thr were also found
R;Hermann, J.; Barel, M.; Frade, R.
Biochem. Biophys. Res. Commun. 204, 453-460, 1994
A;Title: Human erythrocyte ankyrin, a cytoskeleton component, generates the p57 membrane
A;Reference number: PC2220; UID:95071348; PMID:7526850
A;Accession: PC2220
A:Molecule type: protein
A;Residues: 910-929 <HER>
J. Biol. Chem. 265, 10589-10596, 1990
A;Title: Mapping the binding sites of human erythrocyte ankyrin for the anion exchanger
A;Reference number: A35443; UID:90285190; PMID:2141335
A;Accession: A35443
A:Molecule type: protein
A;Residues: 'X','S','X','7-12;403-417,'X',419-422,'H',424,'IQ';797-800,'L',802-814;862-863,

A:Reference number: A35049; MUID:90175370; PMID:1689849

A:Accession: A35049

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1880 <LAW>

A:Cross-references: GB:M2880

C:Genetics:

A:Gene: GDB:ANK1; ANK

A:Cross-references: GDB:118737; OMIM:182900

A:Map position: 8p11.2-8p11.2

C:Superfamily: ankyrin; ankyrin repeat homology

C:Keywords: alternative splicing; cytoskeleton

F:2-1880/Product: ankyrin 1, erythrocyte form 2 #status predicted <MA2>

F:2-1513 1676-1880/Product: ankyrin 2.2, erythrocyte #status predicted <MA2>

F:44-76/Domain: ankyrin repeat homology <AN01>

F:77-109/Domain: ankyrin repeat homology <AN02>

F:110-142/Domain: ankyrin repeat homology <AN03>

F:143-171/Domain: ankyrin repeat homology <AN04>

F:172-204/Domain: ankyrin repeat homology <AN05>

F:205-237/Domain: ankyrin repeat homology <AN06>

F:238-270/Domain: ankyrin repeat homology <AN07>

F:271-303/Domain: ankyrin repeat homology <AN08>

F:304-336/Domain: ankyrin repeat homology <AN09>

F:337-369/Domain: ankyrin repeat homology <AN10>

F:370-402/Domain: ankyrin repeat homology <AN11>

F:403-435/Domain: ankyrin repeat homology <AN12>

F:436-468/Domain: ankyrin repeat homology <AN13>

F:469-501/Domain: ankyrin repeat homology <AN14>

F:502-534/Domain: ankyrin repeat homology <AN15>

F:535-567/Domain: ankyrin repeat homology <AN16>

F:568-600/Domain: ankyrin repeat homology <AN17>

F:601-633/Domain: ankyrin repeat homology <AN18>

F:634-666/Domain: ankyrin repeat homology <AN19>

F:667-699/Domain: ankyrin repeat homology <AN20>

F:700-732/Domain: ankyrin repeat homology <AN21>

F:733-765/Domain: ankyrin repeat homology <AN22>

F:766-798/Domain: ankyrin repeat homology <AN23>

Query Match 9.5%; Score 221.5; DB 2; Length 1880;
Best Local Similarity 27.6%; Pred. No. 1.6e-09;
Matches 79; Conservative 51; Mismatches 115; Indels 41; Gaps 7;

QY 169 ALNQAVYDNDSTLDOLLAQRYKRNRSRGVGTPLRLAASVGHLSCLQVLAHCA 228

Db 176 ALHIAARNDDTRFAAVLLQNDPNPDLV-SKTGP-----TPHIAAHYENLNVAQLLNRS 230

QY 229 DVDSLDVKAQTPLFTAVSHGHLDVRLLEAGASPGSGSYNNCSPVLTAAARDGAVAILQE 288

Db 231 SVNFTPQNGITPLHIAASRGVNTWVLLDRGAQIETKTDELTPHCAARNGHVRIS 290

QY 289 LLDHGAENVKAKLPWASNIASCSGPLYLAAYVGHLDLFRLLHLHGADPDYNTDQGLL 348

Db 291 LLDHGAPIOAKTK-----NGLSPIHMAAQGDHLDVRLLLQYDAIDIDTLDH--- 338

QY 349 ARVPRPTLELCLHNCPEPEYIQLLDIGANILYLSLS-----LDLNS 392

Db 339 ----LTPHVAHCGHH-----RVAKVLLDGKPNRSALNGFTPLHIAKKNHVRVWELL 391

QY 393 QDDKGIALLQARATPRSLLSQV-RLVVRALCAQCPQAINQLDI 437

Db 392 KTCASIDAVTESGLTPLHVASFMGHLPIVKNLLQRCASPNVSNKV 437

RESULT 7

T15347

ankyrin-related unc-44 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Aug-2002

C:Accession: T15346; T15344; T15344; T15344; A57282; B57282; C57282

R:Gatung, S.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid B0350.

A:Reference number: Z18332

A:Accession: T15347

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-2039 <GAT>

A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208873; PIDN:AAA93443.1

A:Accession: T15346

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1000,'SKLOHRT',1002-1718,'KWEELNRL',1727,'D','ES',1942,'PSPAQRS',1950,'IVA

32,'S',2034-2035,'GSPTRRSVPEEHRHSHQEDHEGST' <GA>

A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208875; PIDN:AAA93445.1

A:Accession: T15344

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1718,'KW',1903-1905,'NRLADESPS',1916-1917,'QRSTIVASTSEQVPE',1934-1935,'E'

<GA3>

A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208874; PIDN:AAA93444.1

A:Accession: T15345

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1718,'KWEELNRL',1727,'D','ES',1942,'PSPAQRS',1950,'IVAES',1956-1957,'EQVPE'

PTRRSVPEEHRHSHQEDHEGST' <GA4>

A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208876; PIDN:AAA93446.1

R.Otsuka, A.J.; Franco, R.; Yang, B.; Shim, K.H.; Tang, L.Z.; Zhang, Y.Y.; Boontrakulpoo

J. Cell Biol. 129, 1081-1092, 1995

A:Title: An ankyrin-related gene (unc-44) is necessary for proper axonal guidance in Cae

A:Reference number: A57282; MUID:95263863; PMID:7744957

A:Accession: A57282

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-852,'GGG',856-1000,'SKLOHRT',1002-1319,'IG',1322-1595,'DA',1598-1718,'KWEEL

'SHRED',2007-2008,'TI',2011,2017,'TT',2020-2022,'SHIS' <OTS>

A:Cross-references: GB:U21734; NID:g790607; PIDN:AAA85854.1; PID:g790608

A:Accession: B57282

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 831-852,'GGG',856-1319,'IG',1322-1595,'DA',1598-1718,'KWEELNRL',1727,1799,'E'

V',1945-1947,'VT',1950,'SH',1975,'SESP',1980-1981,'SPTRRSVPEEHRHS',1984-1985,'EDHEGS',1,

A:Cross-references: GB:U21733; NID:g790603; PIDN:AAA85853.1; PID:g790604

A:Accession: C57282

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 194,'F',196,'T',198,'DC',201,'G',409,'AV',412,'O',414,824,'S',826,'ER',829,'E'

4,'TIV',1828,'ESTS',1833,'QV',1836,'E',1934-1935,'EQS',1939,'ESES',1944,'REDDGTIVT',194

A:Cross-references: GB:U21732; NID:g790603; PIDN:AAA85853.1; PID:g790604

C:Genetics:

A:Gene: CESP:unc-44

A:Introns: 97/3; 262/2; 535/1; 749/1; 786/3; 883/3; 1001/2; 1045/3; 1660/1; 1719/2; 1797/

C:Superfamily: ankyrin; ankyrin repeat homology

F:164-192/Domain: ankyrin repeat homology <AN04>

F:358-390/Domain: ankyrin repeat homology <AN1>

F:391-423/Domain: ankyrin repeat homology <AN11>

Query Match 9.3%; Score 218.5; DB 2; Length 2039;

Best Local Similarity 32.2%; Pred. No. 3.1e-09;

Matches 69; Conservative 33; Mismatches 81; Indels 31; Gaps 6;

QY 206 TPLRLAASVGHLSCLQVLAHAGADVDSLDVKAQTPLFTAVSHGHLDVRLVLEAGASPG 265

Db 460 TELHLAARANQTDVVRVLRNGAKVDAQARELOTPHLIASRLGNTDIVILLQAGANSNA 519

QY 266 SYNNCSPVLTAAARDGAVAILQELLDHGAENVKAKLPWASNIASCSG--PLYLAAYVG 323

Db 520 TTRDNYSPHLIAAKEGQEEVAGILLDHNADKTLTK-----KGFTFLHLASKYG 568

QY 324 HLDCFRLLLLHGADPDYNTDQGLLARVPRPTLLEICLHNCPEPEYIQLLDIFGA---- 379

Db 569 NLEVVRLLLERGTPTVIEGKNQ-----VTPHVAHAYN-NDKVAMLLLENGASAKA 618

QY 380 ---NTYLPFSLSLDLTSQDDKGIALLQARATPRS 410

Db 619 AAKNGYTP-LHIAAKNQMEIATSLQFKADPNA 651

```
RESULT 8
T13940
Ankyrin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13940
R:Dubreuil, R.R.; Yu, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 10285-10289, 1994
A:Title: Ankyrin and beta-spectrin accumulate independently of alpha-spectrin in Drosophila
A:Reference number: Z17820; MUID:95024098; PMID:7937942
A:Accession: T13940
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1549 <DUB>
A:Cross-references: EMBL:L35601; NID:G557083; PID:G557084; PIDN:AAC37208.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0011747

Query Match          9.2%; Score 214.5; DB 2; Length 1549;
Best Local Similarity 25.2%; Pred. No. 4.5e-09;
Matches 84; Conservative 53; Mismatches 121; Indels 75; Gaps 10;

QY 131 IMRIVQLAKMN-----LMDITKIFSLQDPKEEDTDEEKQALNQAVY 175
DB 480 IIRILLRSKAVDAIVREGQTPLVHVSRLGNINIIMLLQHGAEINQSNKYSAHIAAK 539
QY 176 DNDSTYLDQLRQERYKRFINSRSGVPGTPLRLAASVGHLSCLQVILLAHGADVDSLDV 235
DB 540 EQENIVQVLENGAENNAV- TKKGF-----TPHLACKYQKQNVQILLQNGASIDFQK 594
QY 236 KAQTPLVAVSHGHLCVVRVLEAGASPGSIYNNCSPLVLTARDGAVAILQELLDHGAE 295
DB 595 NDVTPLVHATHYNNPSIVELLKNGSPNLCAHQCAIHIACKNYLBIAMQLLQHGAD 654
QY 296 ANVKAKLPVWASNIASCSG-----PLYLAAYVGHLDGCFRLLE 332
DB 655 VNIISKSGFPLHLAAGGVNDVQLLEVGIVISAANKGLTEPHVAQEGHVLVQILL 714
QY 333 LHGADPDYNTDQGLLARVPRPTLEICLHNCPEYIQLL-----IDFGANI-Y 382
DB 715 EHGAN-----ISERTNGYTPHMAHYG-----HLDDVKFFIENDADIEMSSNIGY 761
QY 383 LPSLSLDLTSQDDKG-----IALQLQARATPRSL 411
DB 762 TP-----LHQAQOQHIMIINLLRHKAMPNAL 789

RESULT 9
T37958
Probable myotubularin-like protein-tyrosine phosphatase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37958
R:Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21756
A:Accession: T37958
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-559 <SKE>
A:Cross-references: EMBL:Z98974; PIDN:CAB11639.1; GSPDB:GN00066; SPDB:SPAC19A8.03
A:Experimental source: strain 972h-; cosmid c19A8
C:Genetics:
A:Gene: SPDB:SPAC19A8.03
A:Map position: 1
A:Introns: 8/3; 200/3

Query Match          9.1%; Score 212.5; DB 2; Length 559;
Best Local Similarity 39.2%; Pred. No. 1.7e-09;
Matches 51; Conservative 20; Mismatches 46; Indels 13; Gaps 5;

QY 4 SGWKLIDPISDFGRMGIPNR-----NWTIDANRNYEICSTYPTPEIVPKSVTLGTV 55
```

```
DB 119 SSWKSFLLNEVRRMGVGDSTQADGAGGNWRTIKINENYSECHSYPOALAVPASISDSVI 178
QY 56 VGSSKFRSKERPVLVSLYKKNAAICRCSOPLSGF-YTRCVDDDELLLEAISQTN--PGS 112
DB 179 YGCKYRSKRNRPRTUTYLHK-NSFSITRASQPLVIGRQNSAQDEKLVFAIFATSIIPGK 237
QY 113 QPMYVVDTRP 122
DB 238 ENL-IVDARP 246

RESULT 10
T14151
Inv protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14151
R:Yokoyama, T.; Mochizuki, T.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z17888
A:Accession: T14151
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1062 <YOK>
A:Cross-references: EMBL:AF034860; NID:G3560522; PID:G3560523; PIDN:AAC34976.1
C:Genetics:
A:Gene: inv
A:Map position: 4

Query Match          9.0%; Score 212; DB 2; Length 1062;
Best Local Similarity 30.1%; Pred. No. 4.4e-09;
Matches 74; Conservative 40; Mismatches 98; Indels 34; Gaps 9;

QY 139 AKNNLMDITKIFSLQDPKEEDTDEEKQALNQAVYNDSTYLDQLRQERYKRFINSR 198
DB 297 AQSNFAETVKVF--LQHPSVKDDSLGRTSPWAAAGKGNDDVLRMTL-SLKSDIDINMS 353
QY 199 SGWGVPGTPLRLAASVGHLSCLQVILLAHGADVDSLDVKAQTPLEFVAVSHGHLDGCVRLLE 258
DB 354 DKYG--GTALHAAALSGHVSTVKLLDNDDAQVDATDVMKHTPLFRACENGHRRDVIOTLIK 411
QY 259 AGASPGSGSIYNNCSPLVLTARDGAVAILQELLDHGAEANVKAKLPVWASNIASCSG--PL 316
DB 412 GEARVDLVDDQGHSLHWAALGKNADVCQILLENKINPNVQ-----DYAGRTPL 460
QY 317 YLAAYVGHLDGCFRLLELLHGHADPDYNTDQGLLARVPRPTLEICLHNCPEYIQL 373
DB 461 QCAAYGGYINCMAVLMMENADP--NIQDK-----EGRT-----ALHWSCNNGYLDAILK 507
QY 374 LIDFGA 379
DB 508 LLDFAA 513

RESULT 11
S57131
Hypothetical protein YUR110w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J2007
C:Species: Saccharomyces cerevisiae
C>Date: 03-Sep-1995 #sequence_revision 16-Feb-1996 #text_change 19-Apr-2002
C:Accession: S57131; S57133; B33478
R:Ramezani Rad, M.; Kirchrath, L.; Hollenberg, C.P.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57111
A:Accession: S57131
A:Molecule type: DNA
A:Residues: 1-688 <RAM>
A:Cross-references: EMBL:Z49610; NID:G1015823; PIDN:CAA9640.1; PID:G1015824; MIPS:YUR11
R:Rose, M.; Koetter, P.; Entian, K.D.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57132
A:Accession: S57133
```

A:Molecule type: DNA
 A:Residues: 1-688 <ROS>
 A:Cross-references: EMBL:Z49610; NID:g1015823; PIDN:CAA9640.1; PID:g1015824; MIPS:YJR11
 Mol. Kinney, D.M.; Lusty, C.J.
 R. Cell. Biol. 9, 4882-4888, 1989
 A:Title: Arginine restriction induced by delta-N-(phosphonacetyl)-L-ornithine signals in
 A:Reference number: A33478; MUID:90097903; PMID:2689869
 A:Accession: B33478
 A:Molecule type: DNA
 A:Residues: 1-84, 'H', '86-88 <KIN>
 A:Cross-references: GB:M30466; NID:g171311; PIDN:AAA66903.1; PID:g808832
 C:Genetics:
 A:Cross-references: SGD:S0003871
 A:Map position: 10R

Query Match 9.0%; Score 210.5; DB 2; Length 688;
 Best Local Similarity 21.5%; Pred. No. 3.3e-09;
 Matches 84; Conservative 56; Mismatches 99; Indels 151; Gaps 15;

QY 6 WKLDIPISDFGRMGIPNRN---WTITDANRNYEICSTYPPETVVPKSVTLGTWVGSSKF 61
 DB 156 WDIYDIKEPRRQGLDSKDETCFWRSLSTVNEHYEFCPTYESKLFVPRSTSDILLKHSKF 215
 QY 62 RSKERVPLSYLKNNNAICRQSPLSGPYT-RCVDDLELL-----EASISOTNPG 111
 DB 216 RSQKRIPLVTHYHKKATDCNLRSSQPLPGLINQRSIQDEKLWVESFNSFCNKDIRTK-- 273
 QY 112 SQPMYVVDTRPKIWHFVLVIMRIVLQAKNLMIDITKIFSLLPQDKEEDTDTFEEKALN 171
 DB 274 ---HVIDAKPR-----TNAQAQWALG 292
 QY 172 QAYVNDSTYLDQLRQRYKRFINSRGVPGTFLRLAASGYHLSCLQLVLAHGADV 231
 DB 293 GGTEINDVY-----NFFLADNNMGVD-----K 314
 QY 232 SLDVKAQTPLFTAVSHGHLDVVRVLEAGASPGGSYNNCSPLVTAARDGAVAILQELLD 291
 DB 315 SLKLPVTTRFLGIDNIH-----IVSNTAAVMT-----EVIC 346
 QY 292 HGAEANV-----KAKLPVMA---SNIASCSGPLYLAAYVGHLDVCFRLLLLHGADPDY 340
 DB 347 QGGDLNLPLEQNLRISQKSNWLUKNTLILKSVMLKSIIFNHSN-----VLVHCSD-GW 401
 QY 341 NCTDQGLLARVPRPRTLLIEICLHHNCEPEY 370
 DB 402 DRTSQVV-----SLLIEICL-----DPFY 419

RESULT 12
 T30255
 invertin - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30255
 R.Morgan, D.; Turnpenny, L.; Goodship, J.; Dai, W.; Majumder, K.; Matthews, L.; Gardner,
 Nature Genet. 20, 149-156, 1998
 A:Title: Inversin, a novel gene in the vertebrate left-right axis pathway, is partially
 A:Reference number: 220793; MUID:98442650; PMID:9771707
 A:Accession: T30255
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-1062 <MOR>
 A:Cross-references: EMBL:AJ010902; NID:g3757671; PIDN:CAA09388.1; PID:g3757672
 A:Experimental source: strain FBV/N
 C:Genetics:
 A:Gene: invs
 A:Map position: 4

Query Match 8.9%; Score 208; DB 2; Length 1062;
 Best Local Similarity 29.7%; Pred. No. 9.4e-09;
 Matches 73; Conservative 41; Mismatches 98; Indels 34; Gaps 9;

QY 139 AKNMLMDITKIFSLLPQDKEEDTDTTEKQALNAQVYNDSTYLDQLLRQRYKRFINSR 198

DB 297 AOSNFAETVKVF--LQHPVSKDSDQEGRTSFWAAAGKGNDDVLRITML-SLKSDIDINMS 353
 QY 199 SGWGVPGTFLRLAASGYHLSCLQLVLAHGADVSLDKAQTPLFTAVSHGHLDVVRVLE 258
 DB 354 DRYG--GTAHAAAGSHGVSTVKLLLDNQADVDATDMKHTPLFRACEMGHRDVIQTLLK 411
 QY 259 AGASPGGSYNNCSPLVTAARDGAVAILQELLDHGAEANVKAKLPVWASNIASCSG--PL 316
 DB 412 GGARDVLDVDDGSHLSLHWAALGNGADVCCILLENKINPNVQ-----DYAGRTPI 460
 QY 317 YLAAYVGHLDVCFRLLLLHGADPDYNTDQGLLARVPRPRTLLIEICLHHNCEPEY----IOL 373
 DB 461 QSAAYGGINCMVLMENNADP--NIQDK-----EGRT-----ALHWSNNGYLDAILK 507
 QY 374 LIDFGA 379
 DB 508 LLDFAA 513

RESULT 13

T42714

ankyrin 3, splice form 2 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000

C:Accession: T42714

R:Peterson, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Vialamas, M.; Turtzo, L.

J. Cell Biol. 130, 313-330, 1995

A:Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene f

the repeat domain.

A:Reference number: Z22237; MUID:95340633; PMID:7615634

A:Accession: T42714

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1765 <P>

A:Cross-references: EMBL:I40632; NID:g710548; PID:g710551; PIDN:AAB01605.1

A:Experimental source: strain C57BL/6J; kidney

C:Genetics:

A:Gene: Ank3

A:Map position: 10

A:Introns: 1587/1

C:Superfamily: ankyrin; ankyrin repeat homology

C:Keywords: alternative splicing

Query Match 8.8%; Score 206.5; DB 2; Length 1765;
 Best Local Similarity 26.5%; Pred. No. 2.5e-08;
 Matches 75; Conservative 38; Mismatches 93; Indels 77; Gaps 9;

QY 206 TPLRLAASGYHLSCLQLVLAHGADVSLDVKA----- 237
 DB 418 TPIHVAAPFGHVNIVSQLMHGASPTNINRGCTALHWAARSQAQEVVRYLVQDGAQVEA 477

QY 238 -----QTPFTAVSHGHLDVVRVLEAGASPGGSYNNCSPLVTAARDGAVAILQELLDH 292
 DB 478 KAKDDQTPPLHISARLGKADIVQQLLQOGASPNAAATTSVGTPLHAAAREGHEDVAAFLD 537

QY 293 GAEANVKAKLPVWASNIASCSG--PLYLAAYVGHLDVCFRLLLLHGADPDYNTDQGLLAR 350
 DB 538 GASLSITTK-----KGFTPLHVAAYGKLEVASLLLOKASPD--AAAGSGGL---- 582

QY 351 VPRPRTLLIEICLHHNCEPEYIQLLDIFGA-----NIYLPSSLDLTSQDDKGITALL-- 401
 DB 583 -----TPLHVAHYD-NQKVALLLLDQGSAPHAARKNGVTPHIAAKKNQMDIATSLLEY 636

QY 402 -LQADATPRSLLSQVRLVVRALCOAGQFQAINQLDIPMLIS 443
 DB 637 GADANAVTRQGIASVHLAAQE-----GHVDMVSLILS 668

RESULT 14

T42715

ankyrin 3, splice form 3 - mouse

C:Species: Mus musculus (house mouse)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C/Accession: T42715
R/Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Vialamas, M.; Turtzo, I.
J. Cell Biol. 130, 313-330, 1995
A>Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A/Reference number: 222237; MUID:95340633; PMID:7615634
A/Accession: T42715
A/Molecule type: mRNA
A/Residues: 1-1940 <PET>
A/Cross-references: EMBL:L40632; NID:g710548; PID:g710549; PIDN:AAB01604.1
A/Experimental source: strain C57BL/6J; kidney
C/Genetics:
A/Gene: Ank3
A/Map position: 10
A/Introns: 834/1
C/Note: major kidney ankyrin
C/Suprafamily: ankyrin; ankyrin repeat homology
C/Keywords: alternative splicing

Query Match 8.8%; Score 206.5; DB 2; Length 1940;
Best Local Similarity 26.5%; Pred. No. 2.8e-08;
Matches 75; Conservative 38; Mismatches 93; Indels 77; Gaps 9;

QY 206 TPLRLAASYGHLSCLOVLLAHGADVDSLDVKA----- 237
DB 418 TPIHVAAFMGHVNIVSQMLHGHGASPNNTNVRGETALHMAARSGQAEVVRVLYVDGQAVEA 477

QY 238 -----QTPLFTAVSHGHLDCVRVLLLEAGASPGSGSYNNCSYVLTAAARDGAVAILQELLDDH 292
DB 478 KAKDDOTPLHISARLGKADIVQQLQCGASPNNAATTSYGTPLHLAAREGHEDVAAFLDDH 537

QY 293 GAENYKAKLPVWASNIASCSCG--PLYLAAYVGHLDLFRLLLLHGDADPDYNTDQGLLAR 350
DB 538 GASLSITTK-----KGFTPLHVAAYKYLEVASLLQKSASPD--AAGKSGL--- 582

QY 351 VPRPRTLLEICLHNCEPEYIQLLIDFGA-----NIYLPSSLDLTSDODDKGIALL-- 401
DB 583 -----TFLHVAAYHD--NQVALLLLDQGASPHAAKNGYTPHLHAAKKNQMDIATSLLEY 636

QY 402 -LQARATPRLSLLSOVRLVVRALCOAGQPOAINQLDIPPMILIS 443
DB 637 GADANAVTRQGIASVHLAAQE-----GHVDMVSLLS 668

RESULT 15
T42713
ankyrin 3, splice form 1 - mouse
C/Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C/Accession: T42713
R/Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Vialamas, M.; Turtzo, I.
J. Cell Biol. 130, 313-330, 1995
A>Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A/Reference number: 222237; MUID:95340633; PMID:7615634
A/Accession: T42713
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1943 <PET>
A/Cross-references: EMBL:L40632; NID:g710548; PID:g710550; PIDN:AAB01606.1
A/Experimental source: strain C57BL/6J; kidney
C/Genetics:
A/Gene: Ank3
A/Map position: 10
A/Introns: 855/1
C/Function:
A/Description: supposed to play an important role in the polarized distribution of many
A/Note: major kidney ankyrin
C/Suprafamily: ankyrin; ankyrin repeat homology
C/Keywords: alternative splicing

Query Match 8.8%; Score 206.5; DB 2; Length 1943;

Best Local Similarity 26.5%; Pred. No. 2.8e-08;
Matches 75; Conservative 38; Mismatches 93; Indels 77; Gaps 9;

QY 206 TPLRLAASYGHLSCLOVLLAHGADVDSLDVKA----- 237
DB 418 TPIHVAAFMGHVNIVSQMLHGHGASPNNTNVRGETALHMAARSGQAEVVRVLYVDGQAVEA 477

QY 238 -----QTPLFTAVSHGHLDCVRVLLLEAGASPGSGSYNNCSYVLTAAARDGAVAILQELLDDH 292
DB 478 KAKDDOTPLHISARLGKADIVQQLQCGASPNNAATTSYGTPLHLAAREGHEDVAAFLDDH 537

QY 293 GAENYKAKLPVWASNIASCSCG--PLYLAAYVGHLDLFRLLLLHGDADPDYNTDQGLLAR 350
DB 538 GASLSITTK-----KGFTPLHVAAYKYLEVASLLQKSASPD--AAGKSGL--- 582

QY 351 VPRPRTLLEICLHNCEPEYIQLLIDFGA-----NIYLPSSLDLTSDODDKGIALL-- 401
DB 583 -----TFLHVAAYHD--NQVALLLLDQGASPHAAKNGYTPHLHAAKKNQMDIATSLLEY 636

QY 402 -LQARATPRLSLLSOVRLVVRALCOAGQPOAINQLDIPPMILIS 443
DB 637 GADANAVTRQGIASVHLAAQE-----GHVDMVSLLS 668

Search completed: December 1, 2003, 07:39:03
Job time : 48 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:44:17 ; Search time 2290 Seconds
(without alignments)
4765.375 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLDIPDFGRMG.....QAINQLDIPPLMLSYLKHQL 449

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlpl
-MODEL=frame_p2n.model -DB=us09941831/runat_01122003_073527_7841/app_query.fasta_1.647
-Q/cgn2_1/USPTO.spool_p/US09941831/runat_01122003_073527_7841/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09941831 @CGN_1_1_2874 @runat_01122003_073527_7841 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_estba:*
3: em_estin:*
4: em_estin:*
5: em_estov:*
6: em_estov:*
7: em_estro:*
8: em_estro:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estfun:*
17: em_gss_hum:*
18: em_gss_hum:*
19: em_gss_hum:*
20: em_gss_vrt:*
21: em_gss_vrt:*
22: em_gss_vrt:*
23: em_gss_vrt:*
24: em_gss_vrt:*
25: em_gss_vrt:*
26: em_gss_vrt:*
27: em_gss_vrt:*
28: gb_gsl1:*

29: gb_gsl2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1459	62.3	1262	11	AK009644	AK009644 Mus muscu
2	1160	49.5	719	12	BM190670	BM190670 POSM01000
3	1113	47.5	644	13	EX107270	EX107270 BX107270
4	890.5	38.0	861	13	BU254916	BU254916 603747645
5	888	37.9	908	13	BU220575	BU220575 603106716
6	849.5	36.3	678	29	AG127205	AG127205 Pan trogl
7	838	35.8	849	13	BU228932	BU228932 603798270
8	805.5	34.4	713	14	CA364256	CA364256 639132 NC
9	805.5	34.4	758	13	EX073379	EX073379 BX073379
10	776	33.1	446	14	W76256	W76256 zd66f07.r1
11	755.5	32.3	755	14	CA368575	CA368575 644799 NC
12	752	32.1	518	28	AZ725513	AZ725513 RPCI-24-1
13	739.5	31.6	680	14	CA358977	CA358977 631831 NC
14	708.5	30.2	932	14	BY709303	BY709303 BY709303
15	683	29.2	643	10	BB661935	BB661935 BB661935
16	675	28.8	627	10	BB662275	BB662275 BB662275
17	673	28.7	647	14	CB516250	CB516250 ssalrpb50
18	649.5	27.7	990	14	CB207207	CB207207 AGENCOURT
19	647	27.6	1060	29	CC217390	CC217390 CH261-79C
20	640.5	27.3	878	14	CB198136	CB198136 AGENCOURT
21	615.5	26.3	690	13	BU382590	BU382590 603858336
22	608.5	26.0	776	13	BU385175	BU385175 603583214
23	601	25.7	538	12	BI290522	BI290522 UI-R-DK0-
24	585	25.0	482	9	AA633985	AA633985 ac33f04.s
25	577	24.6	1201	9	AL529411	AL529411 AL529411
26	574	24.5	372	28	AZ725227	AZ725227 RPCI-24-1
27	573.5	24.5	675	14	CA375712	CA375712 653860 NC
28	567.5	24.2	524	14	W58534	W58534 zd20c05.r1
29	560.5	23.9	583	13	BU438451	BU438451 604147108
30	544	23.2	655	9	AL776681	AL776681 AL776681
31	534	22.8	486	10	BF889720	BF889720 289285 MA
32	531	22.7	708	14	CB591162	CB591162 AGENCOURT
33	517	22.1	543	14	W72063	W72063 zd66f07.r1
34	516	22.0	567	12	BJ493353	BJ493353 BJ493353
35	480.5	20.5	1301	11	AK006515	AK006515 Mus muscu
36	480	20.5	723	14	CB593182	CB593182 AGENCOURT
37	478	20.4	631	13	BQ395305	BQ395305 NISC Rgl4
38	472	20.1	625	10	AW957331	AW957331 EST369521
39	461	19.7	475	9	AI546016	AI546016 fb77b07.y
40	451	19.2	859	9	AJ454588	AJ454588 AJ454588
41	450	19.2	990	12	BM545864	BM545864 AGENCOURT
42	449	19.2	611	12	BM192381	BM192381 dai22g11.
43	446	19.0	464	14	W58499	W58499 zd20c05.r1
44	446	19.0	561	12	BJ079290	BJ079290 BJ079290
45	443.5	18.9	464	10	BF555515	BF555515 UI-R-A1-d

ALIGNMENTS

RESULT 1
AK009644
LOCUS AK009644 1262 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310036C05 product:ankyrin repeat domain-containing SOCS box protein 12, full insert sequence.
ACCESSION AK009644
VERSION AK009644.1 GI:12844562
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci P. and Hayashizaki Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

2 Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M. and Hayashizaki Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

3 Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A. and Hayashizaki Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

4 Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H., Ashburner M., Batalov S., Casavant T., Fleschmann W., Gaasterland T., Giesi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L. M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M. F., Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M., Guatinchich S., Hill D., Hofmann M., Hume D. A., Kamiya M., Lee N. H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K. F., Suzuki H., Toyooka K., Wang K. H., Weitz C., Whittaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S. and Hayashizaki Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 12102622
 PUBMED 12102622

6 (bases 1 to 1262)
 Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H., Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M., Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori P., Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H., Kawai J., Kojima J., Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T., Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M., Muramatsu M. and Hayashizaki Y.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
 CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATTCGAGTTCGATTAATAATATCCCTCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI, 3' end: SstI.
 Host: SOLR.

Location/Qualifiers
 1..1262
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:2310036C05"
 /db_xref="MGI:1902754"
 /db_xref="taxon:10090"
 /clone="2310036C05"
 /sex="male"
 /tissue type="tongue"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 198..1124
 /note="unnamed protein product; ankyrin repeat domain-containing SOCS box protein 12 (Locustlink|142690, GB|AP403040, evidence: BLASTN, 100%, match=1262) putative"
 /codon_start=1
 /protein_id="BAB26410.1"
 /db_xref="GI:12844563"
 /db_xref="MGI:1917642"
 /translation="MNLMDIAKIFSLLOPEKEEDTGTGEKQALNQAVVNDSCSLDHLHGRYKRFINSRSGMGPCTPLRLAASVGHLCNVKLVLEHGDVLDVKAQTPPLF TAVSHGLECYRVMLEAGACPSGTSYNNCSPLVTSRDSGAFALLOELHGHGAENKVA KLPVWASIASCSGLPLAAYVHLDLCPRLLLYGDADPDYNTDGLLSRVDPQRTLL KLPHNCEPEYIQLIDFGANILPSLPVDPDTSDQDKIGKLLQLQARATPRSLLSQTR LVIRRSLCRANQSQATDQDIPPVLIYLKIQ"

BASE COUNT 318 a 338 c 272 g 334 t

ORIGIN

Alignment Scores:
 Pred. No.: 2,42e-146 Length: 1262
 Score: 1459.00 Matches: 282
 Percent Similarity: 92.94% Conservative: 21
 Best Local Similarity: 86.50% Mismatches: 22
 Query Match: 62.27% Indels: 2
 DB: 11 Gaps: 1

US-09-941-831A-20 (1-449) x AK009644 (1-1262)

QY 124 lleTtpHisPheLeuValIleMetArgIleValLeu---GlnLeuAlaLysMetAen 142
 Db 145 AFTCTGGAATTCCTCTCTCTGCAATGAGATA-GTGCTATTTCGAAATAACCAAGATGAAC 203
 QY 143 LeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAspThr 162
 Db 204 CTCATGGATATCGCCAGATCTCTCTCTCTGCAACCTGAAAGAGAGAGGAGGACACT 263
 QY 163 AspThrGluGluLysGlnAlaLeuAenGlnAlaValTyrAspAenAspSerTyrThrLeu 182
 Db 264 GGCACCGGGGAAAAACAGGCTCTCAATCAAGCTGTATATGACATGACTCTCTGTACCCCTG 323
 QY 183 AspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAenSerArgSerGlyTyrTpGly 202

```

Db      324  GACCACCTTCTACACGAGAACTTATAAAGCGTTTCATCAACGAGGAGTGGCTGGGT 383
Qy      203  ValProGlyThrProLeuArgLeuAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
Db      384  ATACCTGGACACCCCTTGGCTTGGAGCTCTTATGGTCACCTTAATTTGTGGAGGTC 443
Qy      223  LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValIysAlaGlnThrProLeuPhe 242
Db      444  CTCCTGGAACATGGTCTGATGTTGATAGCTTGGATGTCAAAGCACAACACCCTTTTC 503
Qy      243  ThrAlaValSerHisGlyHisLeuSerCysValArgValLeuLeuGluAlaGlyAlaSer 262
Db      504  ACTGCTGTGAGCCACCGCTCATCTGGAGTCGTCGAGAATCTTTTGAAGCTGGTCTGT 563
Qy      263  ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
Db      564  CTTAGTGGTAGATCTACAACTATAGCTTCTGCTCAGTCCCTCATCTGCTGAGTGGGT 623
Qy      283  ValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValIysAlaLeu 302
Db      624  TTTGGCATCTTACAGGAGCTCTAGGCGATGGTGGCGAGGCTAAATGTCAAAGCTAACTA 683
Qy      303  ProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrIleuAlaValTyr 322
Db      684  CCAGTCTGGGCGCTTAAATATAGCTTCTGCGCCCTCTATCTGCTGCAGTCTAT 743
Qy      323  GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCys 342
Db      744  GGGCACCTTGATGTTTCCGCTGCTTTTGCCTCTATGGGCGAGATCCTGATTACAACTGC 803
Qy      343  ThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeu 362
Db      804  ACTGACGAGGCTTTTAAAGTCGTGTTCCAGCGCTCGCACACTCTTGAATCTGCCTT 863
Qy      363  HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyr 382
Db      864  CATCATATTTGTGAGCCAGATACATCCAGCTTTTAAATAGATTTTGGAGCTAACATCTAC 923
Qy      383  LeuProSerLeuSerLeuAspLeuThrSerGlnAspAspIysGlyIleAlaLeuLeu 402
Db      924  CTTTCATCTCTCCCTGTGGACCACTTCCAAAGATGATAAAGGCATCAAAATGCTGCTA 983
Qy      403  GlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAla 422
Db      984  CAAGCCCGAGCCACTCCAGGTCATCTCTGTCCAGACCCGTTAGTTATCCGAGATCC 1043
Qy      423  LeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIle 442
Db      1044  CTCCTCGGGCGAAACCAAGTCACAAGCCAGCCAGCTGATATCCCCCTGTGTGATT 1103
Qy      443  SerTyrLeuIysHisGln 448
Db      1104  AGCTACCTCAAAACATCAA 1121

BM190670      719 bp      mRNA      linear      EST 11-DEC-2001
LOCUS      POSM0100011 E08F porcine skeletal muscle cDNA library (POSM) Sus
DEFINITION      scrofa cDNA 5', mRNA sequence.
ACCESSION      BM190670
VERSION      1
KEYWORDS      EST.
SOURCE      BM190670.1 GI:17526633
ORGANISM      Sus scrofa (pig)
REFERENCE      Sus scrofa
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLES      Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ANALYSIS      Yao, J., Cousins, P., Ernst, C. and Saama, P.
COMMENT      Analysis of expressed sequence tags from a normalized porcine
      skeletal muscle cDNA library
      Unpublished
      Contact: Jianbo Yao

```

Department of Animal Science
Michigan State University
B215 Anthony Hall, East Lansing, MI, USA
Tel: 517-355-8443
Fax: 517-353-1699
Email: yaoj@msu.edu
Seq primer: M13 reverse.

FEATURES

Source
1..719
Location/Qualifiers
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/sex="male and female"
/tissue type="skeletal muscles"
/dev stage="45 d and 90 d of gestation, birth, 7 wk and 1 yr of age"
/lab_host="DH10B"
/clone_lib="porcine skeletal muscle cDNA library (POSM)"
/note="Organ: hind limbs; Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 151 a 206 c 176 g 182 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 1.76e-114 Length: 719
Score: 1160.00 Matches: 222
Percent Similarity: 96.22% Conservative: 7
Best Local Similarity: 93.28% Mismatches: 8
Query Match: 49.51% Indels: 1
DB: 12 Gaps: 0

US-09-941-831a-20 (1-449) x BM190670 (1-719)

```

Qy      133  LeuGlnProAspIysGluGluAspThrAspThrGluGlyLysGlnAlaLeuAsnGln 172
Db      7  GTCCGCGCCGCAAGGAGGAGGAGGACATGACACAGGGGAGAGCAGCCCTCATCA 66
Qy      173  AlaValTyrAspAsnAspSerTyrThrLeuAspGlnIleuLeuArgGlnGluArgTyrLys 192
Db      67  GCAGGTACACACACACTCTATACCTTGGACCACTTTTGGCCAGGAGGATACAA 126
Qy      193  ArgPheIleAsnSerArgSerGly-TrpGlyValProGlyThrProLeuArgLeuAla 212
Db      127  CGATTTCATCAATAGTAGGAGTGGCTTGGGTGTCCTGGACACCTTGGCTTGGCTGC 186
Qy      212  aSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 232
Db      187  TTCCTTATGCCACCTTANCTGCTTGCAGTCTCTCTGCACATGGTGTGAGCTTGACAG 246
Qy      232  rLeuAspValIysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp 252
Db      247  CTTGGACGTCAAGGCACAGACCACTTTTACCCCGCTCAGTCTGTCATGGCCATCTGG 306
Qy      252  sValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys 272
Db      307  TGTGCGTGTGCTTTTGGAGCTGGTGTCTCTCTGTGTGTAGCATCTACCAACTGCTC 366
Qy      272  rProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuAspHis 292
Db      367  TCCTGTGCTCACAGCTGCCCGTATGGTGTCTTGGCCATCTTGAAGAGCTCTTGGGTCA 426
Qy      292  sGlyAlaGluAlaAsnValIysAlaLeuProValTrpAlaSerAsnIleAlaSerCys 312
Db      427  TGGTCAGAGGCCAATCAAGGCAAACTACAGTCTGGTCTCATCAACATAGCTTCATG 486
Qy      312  sSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeu 332
Db      487  TTCTGGCCCCCTCTATTATTGGCTGAGTGTATGGTCACCTTGACTGTTCCGCTGCTTT 546
Qy      332  uLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgVal 352
Db      547  GCTCCACGAGGAGATCCTGACTATAAATGCTACTGATCAACACCTGTGTGGCTGCTGC 606

```



```

352 oArgProArgThrLeuLeuGluLeuCysLeuHisAsnCysGluProGluTyrIleG1 372
|||||
607 ACGGCCCCGACCCCTTAAATTTGCTCCACCAATAATGTGACCCAGAGTATATCCA 666
|||||
372 nLeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeuAsp 389
|||||
667 GCTGTTAATTGATTNGTGTAACTGACCTTACCTTNCATCTCTCCCTGGAC 718
|||||

RESULT 3
BX107270 644 bp mRNA linear EST 06-FEB-2003
LOCUS BX107270 Soares fetal heart NBH19W Homo sapiens cDNA clone
DEFINITION IMAGp998N09778 ; IMAGE:341192, mRNA sequence.
ACCESSION BX107270
VERSION BX107270.1 GI:27847318
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 644)
Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
Radelof, U., Schneider, D. and Korn, B.
Human Unigeneset - RZPD3
Unpublished
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD: IMAGp998N09778.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACAGGAAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..644
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998N09778 ; IMAGE:341192"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBH19W"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."
BASE COUNT 125 a 194 c 147 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 1,73e-109 Length: 644
Score: 113.00 Matches: 214
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 47.50% Indels: 0
DB: 13 Gaps: 0

```

```

US-09-941-831A-20 (1-449) x BX107270 (1-644)
QY 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
|||||
DB 3 CCGTGGACACCCCTTGGCTGTGGCTCTTTATGCGCCACTTGAAGCTGTTTGAAGTCTCTC 62
|||||
QY 224 LeuAlaHisGlyAlaAspValAspSerLeuAspValValAlaGlnThrProLeuPheThr 243
|||||
DB 63 TTAGGCCCATGGTGTGATGTTGACAGCTTGGATGCAAGGCACAGCCACTTTTCTACT 122
|||||
QY 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
|||||
DB 123 GCTGTCACTCATGGCCATCTGGACTGTGTACGTGTGCTTTTGGAGCTGGTCTCTCTCT 182
|||||
QY 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
|||||
DB 183 GGTGTGTAGCATCTACACAACTGTTCTCCGCTGCTCACAGCTGCCCGTGTGCTGTGT 242
|||||
QY 284 AlaTleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValValAlaValLeuPro 303
|||||
DB 243 GCTATCTCTGCAGGAGCTCTTAGACCATGTGTGAGAGGCCAAGCTCAAAGCTAAACTACCA 302
|||||
QY 304 ValTTPAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323
|||||
DB 303 GTCTGGGCATCAACACATAGCTTCATGTTCTGGCCCCCTCTATTGGCCGCTGTACGGG 362
|||||
QY 324 HisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThr 343
|||||
DB 363 CACCTTGGAGCTGTTTCCGCTGCTTTTGTCTCCAGCGGCGAGACCTGTACAACTGCAC 422
|||||
QY 344 AspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeuCysLeuHis 363
|||||
DB 423 GACCAAGGCGCTATTGGCTGTGCTCCAAAGCCCCACCCCTCTTGAATCTTGCCTCCAT 482
|||||
QY 364 HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeu 383
|||||
DB 483 CATTAATTGTGACCCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTCTATATCTACCTT 542
|||||
QY 384 ProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuGln 403
|||||
DB 543 CCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAGGCATTCGATTCGTGTCTACAG 602
|||||
QY 404 AlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeu 417
|||||
DB 603 GCCCGAGGCACTCCACGGTCACTTCTATCACAGGTCCCGTTTA 644
|||||

RESULT 4
BX254916 861 bp mRNA linear EST 26-NOV-2002
LOCUS 603747845F1 CSEQCHN38 Gallus gallus CDNA clone CHEST655111 5', mRNA
DEFINITION sequence.
ACCESSION BU254916
VERSION BU254916.1 GI:25511095
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 861)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE
PUBMED 12445392
CONTACT: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930

```



```
Alignment Scores:
Pred. No.: 5,61e-85 Length: 908
Score: 888.00 Matches: 175
Percent Similarity: 82.38% Conservative: 40
Best Local Similarity: 67.05% Mismatches: 40
Query Match: 37.90% Indels: 8
DB: 13 Gaps: 2

US-09-941-831A-20 (1-449) x BU220575 (1-908)
QY 128 LeuValLeuLeuMetArgIleValLeu-----GlnLeuAlaLysMetAsnLeuMetAsp 145
DB 128 TTGCTGCTTCAATGAAGTGGTACTCAGAAGACAGACACAAAATAATGACCTTAATGGAC 187
QY 146 IleThrIysIlePheSerLeuLeuGlnProAspLysGluGluAspThrAspThrGlu 165
DB 188 ATCAATAAGATTTTCTCCATGCTCCAAACC-----AGAGAAGATGAAGATAATGGA 238
QY 166 GluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 195
DB 239 GAAAGTGAGAGCTGAACCAAGCGGTGCTGAGATGATTACCAAGCTCTTGACAAGCTC 298
QY 186 LeuArgGlnGluArgTyrIysArgPheIleAsnSerArgSerGlyTyrGlyValProGly 205
DB 299 TTGCACCAAGACAGGTACAGAGAGTATCATCAACAGCAGGAGTGGTGGGTGTTCACAGC 358
QY 206 ThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
DB 359 ACCCCACTGCACCTGGCCGCTCCAAAGGCCACCTCAGGAGCCTGGAGGTCTCTCTGTCT 418
QY 226 HisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaVal 245
DB 419 CATGGGGCGAGGTGACACCTGGACCTGGACCTGAAGGACCAACCCCTCTCTTCACAGCAG 478
QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
DB 479 AGTAATGGCCGATTGGAGTGGTGAAGCACTCTGGAGGCAGGAGCTGGTCCCTCTGGC 538
QY 266 SerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIle 285
DB 539 AGCATCTACAACAATTGCTCACCTCTGCTGACGCGCAGCTAGGATGGGAACGTGTGAGATT 598
QY 286 LeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTyr 305
DB 599 CTGCACAGCTCTCTGGAACATGTGGCCGAGACCAATGTCCAAAGAGGTGGCTGAGTGG 658
QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeu 325
DB 659 GCTGCCAATCCACTGCTGTTCTGCTGCTGCTTACTCTGGGCTGCTATGGGCACCTG 718
QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGln 345
DB 719 GAATGCTTTAGGTGTGCTGCTCTATGTGTGACATCCCAACTACCACTGCACTGATGAG 778
QY 346 GlyLeuLeuAlaArgValProArgProArgThrLeu-LeuGluIleCysLeuHisHis 365
DB 779 AGGATGATTGGCGCATCAAGGAGCCCAAGACTCTGCTTGGAGATCTGCTGAGGCACAG 838
QY 365 nCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSe 385
DB 839 -TGCCGGGT-GAGTTCATCAAGCTGTTCTTGAATTTGGAGCCACTGTGTTTTCGCAAA 896
QY 385 x 385
DB 897 C 897

RESULT 6
AG127205/c
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-137P23.F, genomic survey sequence.
ACCESSION AG127205
VERSION AG127205.1 GI:16656370

KEYWORDS
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
SEQUENCING: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. .678
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-137P23.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 174 a 182 c 178 g 143 t
ORIGIN
Alignment Scores:
Pred. No.: 5e-81 Length: 678
Score: 849.50 Matches: 169
Percent Similarity: 90.6% Conservative: 6
Best Local Similarity: 87.5% Mismatches: 13
Query Match: 36.26% Indels: 5
DB: 29 Gaps: 1
US-09-941-831A-20 (1-449) x AG127205 (1-678)
QY 122 ProLysIleTyrPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMet 141
DB 586 CCACAGATCTGGCAATTCCTTGTGCTCATATGAGATAGTCTCCATATGCAAGATG 527
QY 142 AsnLeuMetAspIleThrIysIlePheSerLeuLeuGlnProAspLysGluGluAsp 161
DB 526 AACCTCATGGACATCACCAGATCTTCTCCCTCTGACGCCCAAGAGGAGGAGAC 467
QY 162 ThrAspThrGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThr 181
DB 466 ACTGACACAGAGGAGAGAGCGCTCTCAATCAAGCAGTGTATGACACGACTCTATCT 407
QY 182 LeuAspGlnLeuLeuArgGlnGluArgTyrIysArgPheIleAsnSerArgSerGlyTyr 201
DB 406 TTGGACCAAGCTTTTGGCCAGGAGGCTTACAAACGCTTCATCAACAGCAGGAGTGGCTGG 347
QY 202 GlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeuGln 221
DB 346 GCGGTCTCTGGACACCTTGGCGTGTGGCTCTTTATGAACACTTGGAGCTGTTTGCA 287
QY 222 ValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeu 241
DB 286 GTCTCTTTAGCCCATGGTGTGCTGATGTTTGCAGAGCTTGGATGTCAAGGCACAGACGCACTT 227
```

QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261
 |||||
 Db 226 TTCACCTGCTGACGAGCCATCTGGAGTGTGACGTGCTTTTGAAGCTGTGGC 167
 |||||
 QY 262 SerProGlyGlySerIleTyrAenAenCysSerProValLeuThrAlaAlaArgAspGly 281
 |||||
 Db 166 TGTCTGTGGTGTAGCATCTACAACACTGTTCTCCCGTGTCTCACAGCTGCCCGTGTGTGT 107
 |||||
 QY 282 AlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValIlysAlaLys 301
 |||||
 Db 106 GCTGTGTCTATCTCTGACAGAGCTCCCGAACGTCGTACGTCGGAAGCTTTCGAACGCT 47
 |||||
 QY 302 LeuProValTrpAlaSerAsnIleAlaSerCysSerGly 314
 |||||
 Db 46 -----AGCATCATCAGCGGGATCCGGC 23
 |||||

RESULT 7
 LOCUS BU228932 849 bp mRNA linear EST 26-NOV-2002
 DEFINITION 603798270F1 CSEQCHN23 Gallus gallus cdna clone CHEST765118 5', mRNA
 sequence.
 ACCESSION BU228932
 VERSION BU228932.1 GI:25468283
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 849)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST
)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1. 849
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST765118"
 /dev_stage="22"
 /lab_host="DH108"
 /clone_lib="CSEQCHN23"
 /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; this normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was bluntended, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

BASE COUNT 198 a 224 c 241 g 186 t
 ORIGIN
 Alignment Scores:

Pred. No.: 1.25e-79 Length: 849
 Score: 838.00 Matches: 158
 Percent Similarity: 84.26% Conservative: 40
 Best Local Similarity: 67.23% Mismatches: 34
 Query Match: 35.77% Indels: 3
 DB: 13 Gaps: 1
 US-09-941-831A-20 (1-449) x BU228932 (1-849)
 QY 215 GlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAsp 234
 |||||
 Db 5 GCCCACTCTAGAGAGCTGAGAGTCTCTCTGTCATGGGCGGAGGTGGACGCTGGAC 64
 |||||
 QY 235 ValIysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArg 254
 |||||
 Db 65 GTGAAGGCACAAACCCCTCTCTCACAGCAGTCAGTAATGGCCACTTGGAGTGGTGAAA 124
 |||||
 QY 255 ValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProVal 274
 |||||
 Db 125 GCGCTGCTGGAGGAGGAGCTGCTCCCTCTGCGCAGCATCTACAACAATTCCTCACCATG 184
 |||||
 QY 275 LeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAla 294
 |||||
 Db 185 CTGACGCGCAGCTAGGATGGGAACGTTGAGATTCTGCACAGCTCTCTGAAACATGCTGCC 244
 |||||
 QY 295 GluAlaAsnValIysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGly 314
 |||||
 Db 245 GAGACCAATGTCACAGCAGGAGTGTCTGAGTGGGTGGCCACTCCACTGCTGTGTCGGT 304
 |||||
 QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
 |||||
 Db 305 CCGCTTTACCTCGCGGTGTCTAAGGGCACCTGGAATGCTTTTAGGTGCTGCTCTAT 364
 |||||
 QY 335 GlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
 |||||
 Db 365 GGTGCCGATCCCACTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 424
 |||||
 QY 355 ArgThrLeuLeuGluIleCysLeuHisAsnCysGluProGluTyrIleGlnLeuLeu 374
 |||||
 Db 425 AAGACTCTGCTGGAGATCTGCTGAGGCACACTGCGCGGTGAGTTCATCAAGCTGCTT 484
 |||||
 QY 375 IleAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAsp 394
 |||||
 Db 485 CTGTGACTTTGGAGCAATGTGTATTGTCACCAACATCAAG-----AAGATAGCACCTGGT 538
 |||||
 QY 395 AspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGln 414
 |||||
 Db 539 AGCGAGGGCTAGAGCTGCTGTTCAGGCAAGAGCTCATCCCAAAATCCCTGATGTCTCAA 598
 |||||
 QY 415 ValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGln 434
 |||||
 Db 599 TCTAGGCTGTGTGAGGCACATCTCTGAAGAGGCTGGCTGTGTCACATGCCCTCAGAGAG 658
 |||||
 QY 435 LeuAspIleProPro-MetLeuIleSerTyrLeuLysHisGln 448
 |||||
 Db 659 CTGGACATTCACCACTGCTGCTGAGTACCTCCCAACATCAG 701
 |||||
 RESULT 8
 LOCUS CA364256
 DEFINITION 639132 NCCOWA 1RT Oncorhynchus mykiss cdna clone 1RT86F15_B_C08 5',
 mRNA sequence.
 ACCESSION CA364256
 VERSION CA364256.1 GI:24673292
 KEYWORDS EST.
 SOURCE Oncorhynchus mykiss (rainbow trout)
 ORGANISM Oncorhynchus mykiss
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 1 (bases 1 to 713)
 Rexroad, C.E. and Keele, J.W.
 Sequence analysis of a rainbow trout normalized cDNA library


```
Db 13 AACTACATCGACTGTTCTCATCTGTCAGATTCTTATGATCCAGTTTAAAGAAAGTCTTAGAA 72
Qy 161 AspThrAspThrGluGluysGlnAlaLeuAsnGlnAlaValTyrAspAsnSerTyr 180
Db 73 GAAAAAGACACCGTGAAGTCACTGCTCCATCAGCAGTGTCCACATGATGACAGA 132
Qy 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
Db 133 CTCCTTTGATGAACCTGCTCTCAAGAAAAGTACAGAAAGTTCACTCACTAGGACGGC 192
Qy 201 TTPGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeu 220
Db 193 TGGGGTATCCCAAGAACACCTTTACGAATGGCTCGCTCTAAAGGTCTATCTCAGGTGCTG 252
Qy 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValArgValLeuGluAlaGly 240
Db 253 CAGTTTCTGCTGCCCCATGAGCAGAGTGGATCTGTAGACGTGAAGGCTCAGACCCCA 312
Qy 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
Db 313 CTTTTTCAGACGTCTGTGGGAAATACTTCAACTGTGTTTATGCCCTCCTCAGGCGAGGT 372
Qy 261 AlaserProGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
Db 373 GCTAACCCCAATGCGAGCTTTGTATAACAACACTGCTCTCCGGTCTGACTGCAGCCAGGAA 432
Qy 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaValLeuAla 300
Db 433 GGGGATGTGAGATTGTGAAGCACTTCTTAACATGCTGCTGAGTCAACTCCAGATCT 492
Qy 301 LysLeuProValTpaLaserAsnIleAlaserCysSerGlyProLeuTyrLeuAlaAla 320
Db 493 AAGTCTGTGTTACTTCAAGTCCAGTGTCTTAGTGTGCTCCCTCTATCTGTCTGCT 552
Qy 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
Db 553 GTGTATGGACACTTAGACTGCTTTAAAGTATTACTTCTCTATGGGCTGATCCAAATTC 612
Qy 341 AsnCysThrAspGlnGlyLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
Db 613 AACTGCACAGATCAAAACTACTGAGTAAATCAAGCAGCCTAAGACCGTGTGAATG 672
Qy 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
Db 673 TGCTCAGACATGGCTGTGGGTGGAGTACATCCAACTTTGATTGACTTTGTGTGNCAT 732
Qy 381 IleTyrLeuPro 384
Db 733 GTCTACCTCCCA 744

RESULT 10
W76256 446 bp mRNA linear EST 17-OCT-1996
LOCUS z66f07.r1 Soares fetal heart NBHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:345637 5', similar to PIR:S37771 S37771 ankyrin, erythrocyte -
mouse ;, mRNA sequence.
ACCESSION W76256.1 GI:1386638
VERSION W76256.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 446)
AUTHORS Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marria,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaakis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
JOURNAL Unpublished
COMMENT Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```

```

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 945 Std Error: 0.00
Seq primer: mob.REGA+ET.
FEATURES
Location/Qualifiers
1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1271012"
/db_xref="taxon:9606"
/clone="IMAGE:345637"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBHL19W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
BASE COUNT 78 a 131 c 115 g 122 t
ORIGIN
Alignment Scores:
Pred. No.: 2,17e-73 Length: 446
Score: 776.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.12% Indels: 0
DB: 14 Gaps: 0
US-09-941-831A-20 (1-449) x W76256 (1-446)
Qy 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
Db 2 GTTCTGGGACACCTTGGCTGCTTCTTATGGCCACTTGAGCTGTGTTGCAAGTC 61
Qy 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
Db 62 CTCCTAGCCCATGTTGCTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCACTTTC 121
Qy 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
Db 122 ACTGCTGTGCTAGTCCATCGCCATCTGGACTGTGACGTGTGCTTTTGGAGCTGTGCTCT 181
Qy 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
Db 182 CTGTTGGTGTAGCATCTACAACTGTTCTCCGCTGCTCACAGCTCCCGTGTGTTGCT 241
Qy 283 ValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
Db 242 GTTGTCTCTCTCAGGAGCTCTAGACCATGTTGGAGAGCCCACTCAAGCTAAACTA 301
Qy 303 ProValTpaLaserAsnIleAlaserCysSerGlyProLeuTyrLeuAlaAlaValTyr 322
Db 302 CCAGTCTGGGCATCAACATAGCTTCATGTTCTGGCCCCCTCTATTGGCCGCGAGTCTAC 361
Qy 323 GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCys 342
Db 362 GGGCAGCTTGAGTGTTCGCTGCTTTTGTCTCCAGGGGCGAGCCCTGACTACAACTGC 421
Qy 343 ThrAspGlnGlyLeuLeuAlaArg 350
```

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 945 Std Error: 0.00
Seq primer: mob.REGA+ET.

FEATURES
Location/Qualifiers
1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1271012"
/db_xref="taxon:9606"
/clone="IMAGE:345637"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBHL19W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."

```

Db      422 ACTGACCGGCGCTATTGGCTCGT 445
RESULT 11
CA368575
LOCUS   755 bp mRNA linear EST 06-NOV-2002
DEFINITION 64799 NCCWA lrt Oncorhynchus mykiss cDNA clone lrt154H07_B_D04
5', mRNA sequence.
ACCESSION CA368575
VERSION   CA368575.1 GI:24680140
KEYWORDS  EST.
SOURCE   Oncorhynchus mykiss (rainbow trout)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 755)
Rexroad, C.E. and Keele, J.W.
Sequence analysis of a rainbow trout normalized cDNA library
Unpublished
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Lestown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.
Seq primer: AGCGATACAAATTTCACACGGA.
FEATURES
    source
        1..755
            /organism="Oncorhynchus mykiss"
            /mol_type="mRNA"
            /db_xref="taxon:8022"
            /clone="lrt154H07_B_D04"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /clone_lib="NCCWA lrt"
            /notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
            Library made from pooled tissue from brain, gill, liver,
            spleen, muscle, and kidney."
BASE COUNT 179 a 197 c 224 g 155 t
ORIGIN
Alignment Scores:
Pred. No.: 6.4e-71 Length: 755
Score: 756.50 Matches: 144
Percent Similarity: 82.08% Conservative: 30
Best Local Similarity: 67.92% Mismatches: 37
Query Match: 32.29% Indels: 1
DB: 14 Gaps: 1
US-09-941-831A-20 (1-449) x CA368575 (1-755)
QY      140 LysMetAsnLeuMethApIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 159
Db      120 AGATGAGTCTATGAGCGCTCCAGATCTCTCTGCTTCAGGCC---AAGGAGAG 176
QY      160 GluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSer 179
Db      177 GATGAGGAGCAATGAGCATGCCAGGCTCTGAACGAGCGAGTGCAGTGCAGTGCAGTGC 236
QY      180 TyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSer 199
Db      237 GAGGTGTAGTGAAGTGTGTTCCCGAGAGAGGTACAGAAAGTCTATCAACAGCAGACG 296
QY      200 GlyTrpGlyValProGlyThrProLeuArgLeuAlaSerTyrGlyHisLeuSerCys 219
Db      297 GGCTGGGGATCCCTGTACCCCTCTACGGAGTCCCGCAGCAGCAGCAGCAGTGCAGTGT 356
QY      220 LeuGlnValLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThr 239
Db      357 CTGGAGCTCTCTGTCGACGCGGAGGTGGAGTCTAGATGTGAAGGCCACGACC 416

```

```

QY      240 ProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAla 259
Db      417 CCTCTGTTTACAGCGGTGAGTGGTAAACACCTGGAGACTGTGTGTGCTCTTACTAAGGCT 476
QY      260 GlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArg 279
Db      477 GGAGCCGACCCCAACGCGAGCCCGTACAACAACCTGTTCCTCCCGTGTGACCGCCGCCGC 536
QY      280 AspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLys 299
Db      537 GAGGAGAGCTGGAGGTCTCGAGGAGCTGCTTCAGTTCGAGCGCAGGTCGACGCTCG 596
QY      300 AlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAla 319
Db      597 CCCAAGTCCCTGAGTGGGCTCCAAAGCCACAGCTGCAGGGACCCCTGTACATATCC 656
QY      320 AlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAsp 339
Db      657 GCTGTATATGGACACCTGACCTGTTTAAAGCTGCTGCTCCACGGGCTAATCCTAAC 716
QY      340 TyrAsnCysThrAspGlnGlyLeuLeuAlaArgVal 351
Db      717 TATACTGTAACTCAGAGAAGATGCTGCCAGGATC 752
RESULT 12
LOCUS   518 bp DNA linear GSS 24-JAN-2001
DEFINITION RPCI-24-145H23.TU RPCI-24 Mus musculus genomic clone RPCI-24-145H23
genomic survey sequence.
ACCESSION AZ725513
VERSION   AZ725513.1 GI:12472241
KEYWORDS  GSS.
SOURCE    Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 518)
Tsengaye, G., Geer, K., Kroi, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., De Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 145 row: H Column: 23
Seq primer: SP6
Class: BAC ends.
FEATURES
    Location/Qualifiers
        1..518
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-24-145H23"
            /sex="Male"
            /cell_type="Spleen/Brain"
            /clone_lib="RPCI-24"
            /note="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI;
            RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
            library was cloned in the pTRABAC1 cloning vector at the
            BamHI sites using MboI partially digested male C57BL/6J
            DNA."

```


BASE COUNT 125 a 130 c 134 g 129 t
ORIGIN

Alignment Scores:

Pred. No.: 1,07e-70 Length: 518
Score: 752.00 Matches: 148
Percent Similarity: 91.86% Conservative: 10
Best Local Similarity: 86.05% Mismatches: 14
Query Match: 32.10% Indels: 1
DB: 28 Gaps: 0

US-09-941-831A-20 (1-449) x AZ725513 (1-518)

QY 152 LeuLeuGlnProAspLysGluGluAspThrAspThrGluGluLysGlnAlaLeuAsn 171
DB 2 CTCTGCAACCTGAAGAGGAGGAGGACACTGGCACCAGGGAACAGGCTCTGAAT 61
QY 172 GlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyr 191
DB 62 CAAGCTGTCTATGACAATGACTGCTGTACCTCGACAC-CTTCTACACAGGAACGTTAT 120
QY 192 LysArgPheIleAsnSerArgSerGlyTyrGlyValProGlyThrProLeuArgLeuAla 211
DB 121 AAACGGCTCATCAACAGCACGAGTGGCTGGGGATTCTTGGAAACACCTTGGCAATGGCA 180
QY 212 AlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 231
DB 181 GCTTCCATGGACACTTAATGTGTGAAGGGCTCTCTGAACATGCTGATGTTGAT 240
QY 232 SerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp 251
DB 241 AGCTTGGATGTCAAGACACAAACACACATTTTCACTGCTGTGAGCCACGCTCATCTGGAG 300
QY 252 CysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys 271
DB 301 TGGCTGAGAATGCTTTTAGAAGCTGGTGCCTGTCTAGTGGTAGCATCTACAAATTCG 360
QY 272 SerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAsp 291
DB 361 TCTCTGTTCTCACTGCCTCACGTGATGGGGCTTTTGCCATCTTACAGAGCTCTTAGGG 420
QY 292 HisGlyAlaGluAlaAsnValLysAlaLysLeuProValTyrAlaSerAsnIleAlaSer 311
DB 421 CATGGTGGCAGGCTAATGTCAAGCTAACTACCACTACAGTCTGGGCGTCAATATAGCTTCA 480
QY 312 CysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323
DB 481 TGTCTGGCCCCCTCTATCTGGCTGCAGTCTATGGG 516

RESULT 13

CA358977 680 bp mRNA linear EST 05-NOV-2002
LOCUS 631831 NCCOWA 1RT Oncorhynchus mykiss cDNA clone 1RT5D15_B_08 5',
DEFINITION mRNA sequence.

ACCESSION CA358977 GI:24604164

VERSION CA358977.1

KEYWORDS EST.

SOURCE Oncorhynchus mykiss (rainbow trout)

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 680)

Rexroad, C.E. and Keele, J.W.

Sequence analysis of a rainbow trout normalized cDNA library

Unpublished

Contact: Rexroad CE

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.

Seq primer: AGCGATAACAAATTTACACACAGGA.

FEATURES

source
1..680
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="1RT5D15_B_08"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="NCCOWA 1RT"
/note="Vector: PCMW SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."

BASE COUNT 159 a 185 c 196 g 140 t
ORIGIN

Alignment Scores:

Pred. No.: 3.69e-69 Length: 680
Score: 739.50 Matches: 141
Percent Similarity: 82.76% Conservative: 27
Best Local Similarity: 69.46% Mismatches: 34
Query Match: 31.56% Indels: 1
DB: 14 Gaps: 1

US-09-941-831A-20 (1-449) x CA358977 (1-680)

QY 140 LysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 159
DB 69 AAGATGAGTCTAATGGATGTCTCAAGATCTTCTCTCTCTTCAGCCC---AAGGAAGAG 125
QY 160 GluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSer 179
DB 126 GATGAGGAGCAATGAGCATGCGCAGGCTCTGAACAGGAGTGTAGAGTGTATGACGTA 185
QY 180 TyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSer 199
DB 186 GAGGTCTTACGTGAGCTGTGTCTCCAGGAGAGCTACAGAAAGTCTATCAACAGCAGAGC 245
QY 200 GlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCys 219
DB 246 GGTCTGGGGATCCCTGTATTACCCCTCTACGAGCTGCCGAGCACACGACACCTGAGGTGT 305
QY 220 LeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThr 239
DB 306 CTGGAGCTCTCTGGAGCAGGAGCGGAGTGGACAGTCTAGATGTGAAGGCCAGACC 365
QY 240 ProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAla 259
DB 366 CTTCTGTTTACAGCGTCAAGTGTAAACACCTGGAGCTGTGTGCTTCTTACTGAGGCT 425
QY 260 GlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArg 279
DB 426 GGAGCCGACCCCAACCGCAGCCCGTACAAACACTGTTCCTCCGCTGCTGACCGCCGCGC 485
QY 280 AspGlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLys 299
DB 486 GAGGAGAGCTGTGAGTCTCTCAGGGAGCTGTTCAGTTCGAGGCCAGGTTCGACGCTGG 545
QY 300 AlaLysLeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAla 319
DB 546 CCCAAGTCCCTGAGTGGGCTTCCACGCCACAGCTTGAGGGGACCCCTGTACATATCC 605
QY 320 AlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAsp 339
DB 606 GCTGTATATGGACACCTGGACTGTTTTTAAGTCTGCTGCTCCACGGGGCTAATCTTAA 665
QY 340 TyrAsnCys 342
DB 666 TATAACTGT 674

RESULT 14

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 1, 2003, 08:00:27 ; Search time 510 Seconds
(without alignments)
9940.284 Million cell updates/sec

Title: US-09-941-831A-6
Perfect score: 1878
Sequence: 1 catgattacgccaagcttgg.....taaaaaaaaaaaaaaaaaa 1878

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1878	100.0	1878	22	AAH46857 Human serine/threose
2	1878	100.0	1878	22	AAH78730 Human HIBCJ89 seri
3	1878	100.0	1878	22	AAS33372 DNA encoding human
C 4	1877.2	100.0	2530	22	AAS33264 DNA encoding human
5	1117.8	59.5	1222	24	ABK11107 Human polynucleoti
6	989.8	52.7	1127	24	ABZ11533 DNA encoding human
7	931	49.6	1069	24	ABK11106 DNA encoding novel
C 8	709.8	37.8	8214	23	AAS77257

9	703.4	37.5	1345	24	ABN60012 Novel human coding
10	241.8	12.9	2353	22	AAF63581 Human phosphatase
11	181.4	9.7	3047	23	ABL20003 Drosophila melanog
12	179.8	9.6	2786	23	ABL06295 Drosophila melanog
C 13	169	9.0	1828	22	ABA19208 Human nervous syst
C 14	169	9.0	10828	22	AAK70743 Human immune/haema
C 15	149.6	8.0	3169	23	ABL17297 Drosophila melanog
C 16	141	7.5	408	23	ABV61004 Human prostate exp
17	135.8	7.2	2482	24	ABQ93336 Human CDNA SQ ID
18	108.8	5.8	2019	19	AAV38672 Mus musculus SOCS7
19	103.6	5.5	3382	24	ABK83507 Human CDNA differe
C 20	100.8	5.4	1266	22	AAS33071 DNA encoding human
C 21	100.8	5.4	1433	23	ABV23231 Human prostate exp
22	100.8	5.4	1433	23	ABV29074 Human prostate exp
23	100.8	5.4	1509	25	AAD49617 Human cytoskeleton
24	100.8	5.4	1540	22	AAS33289 DNA encoding human
25	91	4.8	545	23	AAS73111 DNA encoding novel
C 26	85.4	4.5	220	22	ABA13268 Human nervous syst
27	85.4	4.5	2575	23	AAS73882 DNA encoding novel
28	72.8	3.9	1966	22	AAH14752 Human CDNA sequenc
29	72.8	3.9	2710	22	AAKS2009 Human polynucleoti
30	72.8	3.9	4330	24	AAD40741 Human kinase and p
31	71.4	3.8	550	21	AAC94977 Cat flea hindgut a
32	71.2	3.8	389	24	ABL82288 Homo sapiens SOCS7
33	71.2	3.8	419	19	AAV38673 Human polynucleoti
34	71.2	3.8	2346	22	AAKS2993 Human CDNA clone (
35	70	3.7	802	22	AAH07276 Human ovarian anti
36	69.6	3.7	1024	24	ABQ54306 cDNA encoding huma
37	69.4	3.7	401	21	AAA77749 Colon tumour relat
38	69.4	3.7	401	21	AAI28487 Human colon tumour
39	69.4	3.7	401	25	ABZ32673 Human tyrosine pho
40	67.2	3.6	3411	18	AAI91998 CviJI coding seque
41	66.6	3.5	5496	15	AAQ73395 Drosophila melanog
C 42	65.8	3.5	8397	23	ABL20002 Drosophila melanog
C 43	65.8	3.5	9648	23	ABL06294 Drosophila melanog
C 44	65.8	3.5	9648	23	ABL17296 Human OREF ORF1001
45	64.4	3.4	1434	21	AAC75446

ALIGNMENTS

RESULT 1

AAH46857

ID AAH46857 standard; cDNA; 1878 BP.

XX AAH46857;

AC AAH46857;

DT 25-SEP-2001 (first entry)

DE Human serine/threonine phosphatase encoding cDNA (clone ID HIBCJ89).

XX Serine/threonine phosphatase; nootropic; neuroprotective; cytostatic;

KW dermatological; immunosuppressive; antiinflammatory; antibacterial; ss;

KW anti-HIV; antiparkinsonian; antisticking; antianemic; antiarthritic;

KW antirheumatic; virucide; hepatotropic; cerebroprotective; vulnerary;

KW antiinflammatory; nephrotropic; gene therapy; vaccine.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 367..1716

FT /*tag= a

XX WO200155388-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01395.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 02-MAR-2000; 2000US-0186350.

```
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PA
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI; 2001-476208/51.
XX DR P-PSDB; AAB85477.
XX
XX PT Novel proteins of serine/threonine phosphatase family, useful for
XX PT diagnosing, treating, preventing and/or prognosing disorders related to
XX PT the proteins, including cancer, immune response and neuronal disorders
XX PT
XX PS Claim 1; Page 426; 441pp; English.
XX
XX CC The invention provides serine/threonine phosphatase family polypeptides
XX CC and polynucleotides encoding them. The polypeptides can be expressed by
XX CC standard recombinant methodology. The polypeptides, polynucleotides and
XX CC antibodies are useful for diagnosis, prognosis, prevention and treatment
XX CC of neurodegenerative disorders, immune system disorders, autoimmune
XX CC diseases, allergic reactions, infectious diseases, hyperproliferative
XX CC disorders, renal disorders, cardiovascular disorders, cerebrovascular
XX CC disorders, respiratory disorders, endocrine disorders, gastrointestinal
XX CC disorders and also muscular, reproductive disorders (see AAH46846 for a
XX CC detailed description of the diseases that can be treated). They are also
XX CC useful as immune system enhancers, immunosuppressive agents, stimulator
XX CC of B-cell responsiveness to pathogens, activator of T-cells, to induce
XX CC higher affinity antibodies, as an agent to increase serum immunoglobulin
XX CC concentrations, to accelerate recovery of immunocompromised individuals,
XX CC to boost immunoresponsiveness in aged populations and/or neonates, as
XX CC regulator of antigen presentation, as a means to induce tumour
XX CC proliferation in pathologies e.g. AIDS, and/or common variable immuno
XX CC deficiency. The polypeptides and polynucleotides are useful to prevent
XX CC skin aging, for preventing hair loss, to stimulate growth and
XX CC differentiation of hemopoietic cells and bone marrow cells, for
XX CC supporting cell culture of primary tissues, to modulate mammalian
XX CC characteristics such as body weight, height, eye color, hair color and
XX CC skin, to modulate mammalian metabolism to change a mammal's mental or
XX CC physical state, and as food additive or preservative. The present
XX CC sequence represents a human serine/threonine phosphatase encoding cDNA.
XX
XX SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Query Match 100.0%; Score 1878; DB 22; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGATTAGCCCAAGCTTGGCAGGAGGTAGAAAACGTGAAATTTGGTGGATCGTTATGTG 60
DB |||||
DB 1 CATGATTAGCCCAAGCTTGGCAGGAGGTAGAAAACGTGAAATTTGGTGGATCGTTATGTG 60
QY 61 AGTAAGAAACCAAGTAAATGGATTTCTTTATCTTGAACCAACCACTGATCTATGTGGAG 120
DB |||||
DB 61 AGTAAGAAACCAAGTAAATGGATTTCTTTATCTTGAACCAACCACTGATCTATGTGGAG 120
QY 121 GCTTCAGGTGCAGCCCGGAAAGAAACATGATGTCACATCCATCATTGCCACTGTGGAG 180
DB |||||
DB 121 GCTTCAGGTGCAGCCCGGAAAGAAACATGATGTCACATCCATCATTGCCACTGTGGAG 180
QY 181 AAGTTACCACCTACTAGCTGGGTCTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTG 240
DB |||||
DB 181 AAGTTACCACCTACTAGCTGGGTCTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTG 240
QY 241 GCCCACTTTGTTTAGATTTCTGACCTTGTGTGCATGAGTTTATATTTTCACTGCTCAAG 300
DB |||||
DB 241 GCCCACTTTGTTTAGATTTCTGACCTTGTGTGCATGAGTTTATATTTTCACTGCTCAAG 300
QY 301 CTTTCTCAGCCAGCATTAACCTGAAGATCTTTATGCTTTTCTTATATCCCAATCCTCA 360
DB |||||
DB 301 CTTTCTCAGCCAGCATTAACCTGAAGATCTTTATGCTTTTCTTATATCCCAATCCTCA 360
QY 361 AAAGAGATGAGGGAAGTGGATGGAAATCTGATTGACCCCAATATCAGACTTTGGCGGTATG 420
DB |||||

361 AAAGAGATGAGGGAAGTGGATGGAAATCTGATTGACCCCAATATCAGACTTTGGCGGTATG 420
421 GGAATATCCCAACAGAAAACCTGGACCAATACAGATGCAACAGAAAACCTATGAGATATGCAGC 480
421 GGAATATCCCAACAGAAAACCTGGACCAATACAGATGCAACAGAAAACCTATGAGATATGCAGC 480
481 ACCTACCTCTCTGAAATAGTGGTTCCTAAATCTGTTTACCTTGGGAAACGGTGTGGAAGT 540
481 ACCTACCTCTCTGAAATAGTGGTTCCTAAATCTGTTTACCTTGGGAAACGGTGTGGAAGT 540
541 TCAAAAGTTTCAAGAGTAAAGAACGTCCTCTGTCCTCTCTACCTCTCTCAAAAGAGAACAAT 600
541 TCAAAAGTTTCAAGAGTAAAGAACGTCCTCTGTCCTCTCTACCTCTCTCAAAAGAGAACAAT 600
601 GGTGCATTTGGCGCTGTAGCCAGCTCTCTCTGATTTTACACTCGCTGTGTAGATGAT 660
601 GGTGCATTTGGCGCTGTAGCCAGCTCTCTCTGATTTTACACTCGCTGTGTAGATGAT 660
661 GAGCTCTTGTGGAGGCCATTAGCCAAAACCAACCAAGGAGCCAGTTTATGTATGTTGTA 720
661 GAGCTCTTGTGGAGGCCATTAGCCAAAACCAACCAAGGAGCCAGTTTATGTATGTTGTA 720
721 GACACAAGACCAAAAGATCTGGCAATTTCTTGTGCTCATAATGAGAATAGTTCTCCAATTA 780
721 GACACAAGACCAAAAGATCTGGCAATTTCTTGTGCTCATAATGAGAATAGTTCTCCAATTA 780
781 GCCAAGATGAACCTCATGGAATACCAAGATCTTCTCTCTCTGAGCCCAAGGAG 840
781 GCCAAGATGAACCTCATGGAATACCAAGATCTTCTCTCTCTGAGCCCAAGGAG 840
841 GAGGAGGACACTGACACAGAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACAAACGAC 900
841 GAGGAGGACACTGACACAGAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACAAACGAC 900
901 TCCTATACTTTGGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAG 960
901 TCCTATACTTTGGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAG 960
961 AGTGCCTGGGTGTTCTTGGGACACCTTGGCTTGGCTGCTTCTTATGGCCACTTGAGC 1020
961 AGTGCCTGGGTGTTCTTGGGACACCTTGGCTTGGCTGCTTCTTATGGCCACTTGAGC 1020
1021 TGTTTTGCAAGTCTCTTATAGCCCATGCTGATGTTGACAGCTTGGATGTCAAGGCACAG 1080
1021 TGTTTTGCAAGTCTCTTATAGCCCATGCTGATGTTGACAGCTTGGATGTCAAGGCACAG 1080
1081 AGCCCACTTTTTCACCTGCTCACTGATCATGGCCATCTGGACCTGTGTACGTGTCTTTTGGAA 1140
1081 AGCCCACTTTTTCACCTGCTCACTGATCATGGCCATCTGGACCTGTGTACGTGTCTTTTGGAA 1140
1141 GCTGTGCTCTCTCTGTTGGTAGATCTTACAACTGTTCTTCCCGTGTCTCAAGTGGCC 1200
1141 GCTGTGCTCTCTCTGTTGGTAGATCTTACAACTGTTCTTCCCGTGTCTCAAGTGGCC 1200
1201 CGTGTGCTGTGTGCTATCTCTGAGGAGCTCTCTAGACCATGCTGGTGGAGGCCAACGTC 1260
1201 CGTGTGCTGTGTGCTATCTCTGAGGAGCTCTCTAGACCATGCTGGTGGAGGCCAACGTC 1260
1261 AAAGCTAAACTACCACTCTGGGCATCAAACTAGCTTCTATGTTCTTGGCCCCCTCTATTG 1320
1261 AAAGCTAAACTACCACTCTGGGCATCAAACTAGCTTCTATGTTCTTGGCCCCCTCTATTG 1320
1321 GCCGAGCTCTACGGGACCTTGGACCTGTTTCCGCTGCTTTTGTCTCAAGGGGAGACCCCT 1380
1321 GCCGAGCTCTACGGGACCTTGGACCTGTTTCCGCTGCTTTTGTCTCAAGGGGAGACCCCT 1380
1381 GACTACACTGACCTGACCGAGCCCTATTGGCTCTGTGCTCCCAAGACCCCGACCTCTCT 1440
1381 GACTACACTGACCTGACCGAGCCCTATTGGCTCTGTGCTCCCAAGACCCCGACCTCTCT 1440
1441 GAAATCTGCTCCATCAATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTGTT 1500
1441 GAAATCTGCTCCATCAATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTGTT 1500
```

QY 1501 GCTAATATCTACCTTCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCAT 1560
DB |||||
QY 1501 GCTAATATCTACCTTCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCAT 1560
DB |||||
QY 1561 GCATTGCTGCTACAGCGCCGAGCCACTCCACGGTCACTTCTATCAAGAGTCGGTTAGTC 1620
DB |||||
QY 1561 GCATTGCTGCTACAGCGCCGAGCCACTCCACGGTCACTTCTATCAAGAGTCGGTTAGTC 1620
DB |||||
QY 1621 GTCGCCAGAGCTGTGCGCAGCTGCGCCAGCCACAGCCATCAACAGCTGATATCTCT 1680
DB |||||
QY 1621 GTCGCCAGAGCTGTGCGCAGCTGCGCCAGCCACAGCCATCAACAGCTGATATCTCT 1680
DB |||||
QY 1681 CCCATGTTGATTAGCTTACCTAAACACCACTGTATCTTGCAGTCTCCCGAGGAACCTTA 1740
DB |||||
QY 1681 CCCATGTTGATTAGCTTACCTAAACACCACTGTATCTTGCAGTCTCCCGAGGAACCTTA 1740
DB |||||
QY 1741 TGATGCTCCGAAACACCTGCGGACTCACGTAGCTGAGAGCAATTAAGCTCATCCA 1800
DB |||||
QY 1741 TGATGCTCCGAAACACCTGCGGACTCACGTAGCTGAGAGCAATTAAGCTCATCCA 1800
DB |||||
QY 1801 CTTACTGAGCTGCTCTCTCTGATTATCTCCACAAATAAATCTCCAGAAATAAGTA 1860
DB |||||
QY 1801 CTTACTGAGCTGCTCTCTCTGATTATCTCCACAAATAAATCTCCAGAAATAAGTA 1860
DB |||||
QY 1861 AAAAAAAAAAAAAAAAAA 1878
DB |||||
QY 1861 AAAAAAAAAAAAAAAAAA 1878
DB |||||

RESULT 2
AAH78730
ID AAH78730 standard; cDNA; 1878 BP.
XX
AC AAH78730;
XX
DT 03-DEC-2001 (first entry)
XX
DE Human HIBCJ89 serine/threonine phosphatase cDNA sequence.
XX
KW Human; HIBCJ89; ss; serine/threonine phosphatase; PSpase; vaccine;
KW gene therapy; PSpase expression; PSpase expression; PSpase modulation;
KW immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome;
KW Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis;
KW inflammation; Crohn's disease; inflammatory bowel disease; appendicitis;
KW rheumatoid arthritis; cellular proliferative disorder; lymphoma;
KW lung cancer; intestinal cancer; cardiovascular disorder; aneurysm;
KW Scimitar syndrome; Ebstein's anomaly.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 367..1716
FT /*tag= a
FT /product= "HIBCJ89 PSpase protein"
XX
PN WO200164703-A1.
XX
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06256.
XX
PR 02-MAR-2000; 2000US-0186350.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ebner R, Ruben SM;
XX
XX WPI; 2001-530113/58.
DR P-PSDB; AAG77803.
XX
XX Nucleic acids encoding serine/threonine phosphatase polypeptides,
PT useful for preventing, diagnosing and/or treating, e.g. Crohn's

disease, lung cancer and Scimitar syndrome -
Claim 1; Page 308; 335pp; English.

The present sequence represents the specifically claimed human cDNA clone HIBCJ89 which shares homology with members of the serine/threonine family of phosphatases. The invention comprises novel human serine/threonine phosphatase (PSpase) polypeptides and polynucleotides which may be used in the prevention (vaccine), diagnosis and treatment (gene therapy) of disease associated with inappropriate PSpase expression. The PSpase polynucleotides of the invention may be used as DNA probes to detect and quantitate the presence of similar nucleic acids in samples. The PSpase polypeptides may be used as antigens in the production of antibodies against the PSpase polypeptides and in assays to identify modulators of PSpase expression and activity. The anti-PSpase antibodies and antagonists may also be used to down regulate expression and activity, the anti-PSpase antibodies may also be used as diagnostic agents for detecting the presence of PSpase polypeptides in samples. Disorders that may be prevented, diagnosed and/or treated by the invention are: immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome, Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory bowel disease, appendicitis and rheumatoid arthritis); cellular proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and aneurysm).

Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Query Match 100.0%; Score 1878; DB 22; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGATTACGCCAAGCTTGGCAGAGGGTAGAAAACGTAAGTGGTGGATCGTTATG 60
DB |||||
QY 1 CATGATTACGCCAAGCTTGGCAGAGGGTAGAAAACGTAAGTGGTGGATCGTTATG 60
DB |||||
QY 61 AGTAAGAAACACAGCTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGGAG 120
DB |||||
QY 61 AGTAAGAAACACAGCTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGGAG 120
DB |||||
QY 121 GCTTCAGGTGACGCCGCGGAAAGAAACATGATGCACTCCATCATTGCCACTGTGGAG 180
DB |||||
QY 121 GCTTCAGGTGACGCCGCGGAAAGAAACATGATGCACTCCATCATTGCCACTGTGGAG 180
DB |||||
QY 181 AAGTTACCATCATTAGCTGGGTTGTCCTGACCTCCGCTGCAAGAAATTCGGGTG 240
DB |||||
QY 181 AAGTTACCATCATTAGCTGGGTTGTCCTGACCTCCGCTGCAAGAAATTCGGGTG 240
DB |||||
QY 241 GCCCACTTTGTTTATAGATTCTGACCTGTGTGCCATGAGTTTATATTTTCACTGCTCAAG 300
DB |||||
QY 241 GCCCACTTTGTTTATAGATTCTGACCTGTGTGCCATGAGTTTATATTTTCACTGCTCAAG 300
DB |||||
QY 301 CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTTTCTTATAATCCCAATCCTCA 360
DB |||||
QY 301 CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTTTCTTATAATCCCAATCCTCA 360
DB |||||
QY 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATGACCCCAATATCAGACTTTGGCGGTATG 420
DB |||||
QY 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATGACCCCAATATCAGACTTTGGCGGTATG 420
DB |||||
QY 421 GGAATACCCACAGAAACCTGGACCATACAGATGCCAAGAACTATCAGATATGCAGC 480
DB |||||
QY 421 GGAATACCCACAGAAACCTGGACCATACAGATGCCAAGAACTATCAGATATGCAGC 480
DB |||||
QY 481 ACCTACCTCTGAAATAGTGGTTCTTAATCTGTTTACCTGGGAACGGTGGTGAAGT 540
DB |||||
QY 481 ACCTACCTCTGAAATAGTGGTTCTTAATCTGTTTACCTGGGAACGGTGGTGAAGT 540
DB |||||
QY 541 TCAAGATTGAGAAAGTAAAGAAAGTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
DB |||||
QY 541 TCAAGATTGAGAAAGTAAAGAAAGTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
DB |||||

(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (i) and (ii), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAS33043-AAS33486 represent human secreted protein coding sequences, PCR primers, and related sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published pct sequences.

SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Query Match 100.0%; Score 1878; DB 22; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

	Qy	1	CATGATTACGCCCAAGCTTTGGCAGAGGGTAGAAAACGTGAATAAATGGTGCATCGTTATGTG	60
	Db	1		
	Qy	61	AGTAAGAAACCAGCTAAATGGGATCTTTATCTTACTGCAACCCACCTCATCTATCTGTGGAG	120
	Db	61		
	Qy	121	GCTTCAGGTGCAGCGCGGAAAGAACAATGGATTGGACCTCCCATCACATTCGCCACTGTGGAG	180
	Db	121	GCTTCAGGTGCAGCGCGGAAAGAACAATGGATTGGACCTCCCATCACATTCGCCACTGTGGAG	180
	Qy	181	AAGTTACCCATCACTAGCCTGGTGTCCTCGACCTCCCGTGCAGAAATTTCCGGGTG	240
	Db	181	AAGTTACCCCATCTAGCCTGGTGTCCTCGACCTCCCGTGCAGAAATTTCCGGGTG	240
	Qy	241	GCCCACTTTGTTTTAGATTCTGACCTCTGTGTCCTAGAGGTTTATATTTCACTGCTCAAG	300
	Db	241		
	Qy	301	CTTTCTCAGCCAGCATTAACCTGAAGATCTTTATGCTTTTCTTATAATCCCCAAATCCCTCA	360
	Db	301	CTTTCTCAGCCAGCATTAACCTGAAGATCTTTATGCTTTTCTTATAATCCCCAAATCCCTCA	360
	Qy	361	AAAGAGATGAGGGAAAGTGGATGGAAAACTGATTGACCCAATATCAGACTTTGGGCGTATG	420
	Db	361	AAAGAGATGAGGGAAAGTGGATGGAAAACTGATTGACCCAATATCAGACTTTGGGCGTATG	420
	Qy	421	GGATATCCCAACAGAAACTGGACCATAACAGATGCCAACAGAAAATATGAGATATGCACG	480
	Db	421		
	Qy	481	ACCTACCCCTCTGAAATAGTGGTTCTTAAATCTGTTTACCTCGGAAACGGTGGTTGGAAGT	540
	Db	481	ACCTACCCCTCTGAAATAGTGGTTCTTAAATCTGTTTACCTCGGAAACGGTGGTTGGAAGT	540
	Qy	541	TCAAAGTTTCAGAACTGATAAGAAAGCTGCTGCTGCTCTCTACTCTACAAGAGAACAAAT	600
	Db	541	TCAAAGTTTCAGAACTGATAAGAAAGCTGCTGCTGCTCTCTACTCTACAAGAGAACAAAT	600
	Qy	601	GCTGCCATTTCCGCTGTAGCCAGCCTCTCTGGATTTTACACTCGCTGTGTAGATGAT	660
	Db	601	GCTGCCATTTCCGCTGTAGCCAGCCTCTCTGGATTTTACACTCGCTGTGTAGATGAT	660
	Qy	661	GAGCTCTTGTGGAGGCCATTAGCCAAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTA	720
	Db	661	GAGCTCTTGTGGAGGCCATTAGCCAAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTA	720
	Qy	721	GACACAAGACCAAGATCTGGCATTTCCTTGTGCTCATTAATGAGAAATAGTCTTCCAAATTA	780
	Db	721	GACACAAGACCAAGATCTGGCATTTCCTTGTGCTCATTAATGAGAAATAGTCTTCCAAATTA	780

QY 1861 AAAAAAAAAAAAAAAAAA 1878
 DB 1861 AAAAAAAAAAAAAAAAAA 1878
 RESULT 4
 ID AAS33264/c
 XX AAS33264 standard; cDNA; 2530 BP.
 AC AAS33264;
 XX
 DT 04-DEC-2001 (first entry)
 DE DNA encoding human secreted protein, Seq ID No 223.
 XX
 KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
 KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
 KW cerebroprotective; thrombotic; antimicrobial; ophthalmological;
 KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
 KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
 KW Gaucher's disease; neurological disease; cerebrovascular disorder;
 KW thrombosis; wound healing; ss.
 XX
 OS Homo sapiens.
 PN WO200155326-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01347.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-451931/48.
 DR P-PSDB; AAU20555.
 XX
 PT New nucleic acids and polypeptides, useful for diagnosing, preventing
 PT or treating medical conditions -
 XX
 PS Claim 1; SEQ ID No 223; 753pp; English.
 XX
 CC The invention relates to novel isolated nucleic acid molecules (I)
 CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain organs before transplantation, and support cell culture
 CC of primary tissues. AAS33043-AAS33486 represent human secreted protein

CC coding sequences, PCR primers, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2530 BP; 630 A; 602 C; 675 G; 614 T; 9 other;
 Query Match 100.0%; Score 1877.2; DB 22; Length 2530;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1876; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CATGATTACGCCCAAGCTTGGCACGAGGGGTAGAAAAACGTGAAATTTGGTGGATCGTTATGTG 60
 DB 1941 CATGATTACGCCCAAGCTTGGCACGAGGGGTAGAAAAACGTGAAATTTGGTGGATCGTTATGTG 1882
 QY 61 AGTAAGAAACACAGCTTAATGGGATTTCTTATCTTACTCTGCAACCCACCTGATCTATGTGGAG 120
 DB 1881 AGTAAGAAACACAGCTTAATGGGATTTCTTATCTTACTCTGCAACCCACCTGATCTATGTGGAG 1822
 QY 121 GCTTCAGGTGACGCGCGGAAAGAAACATGGATTGCACTCCATCACAATTCGCCACTGTGGAG 180
 DB 1821 GCTTCAGGTGACGCGCGGAAAGAAACATGGATTGCACTCCATCACAATTCGCCACTGTGGAG 1762
 QY 181 AAGTTACCCATCACTAGCTGGGTTGTCCCTGACCCCTCCGCTGCAAGAAATTTCCGGGTG 240
 DB 1761 AAGTTACCCATCACTAGCTGGGTTGTCCCTGACCCCTCCGCTGCAAGAAATTTCCGGGTG 1702
 QY 241 GCCCACTTTGTTTATAGATTCGACCTGTGTGCCATGAGGTTTATATTTCACTGCTCAAG 300
 DB 1701 GCCCACTTTGTTTATAGATTCGACCTGTGTGCCATGAGGTTTATATTTCACTGCTCAAG 1642
 QY 301 CTTTCTCAGCCAGCAATTACCTGAAGATCTTTATGCTTTTCTTATAATCCCAATCCTCA 360
 DB 1641 CTTTCTCAGCCAGCAATTACCTGAAGATCTTTATGCTTTTCTTATAATCCCAATCCTCA 1582
 QY 361 AAAGAGATGAGGGAAGTGGATGGAAACTGATTGACCAATATCAGACTTTTGGGGGTATG 420
 DB 1581 AAAGAGATGAGGGAAGTGGATGGAAACTGATTGACCAATATCAGACTTTTGGGGGTATG 1522
 QY 421 GGAATACCAACAGAAACTGGACCAATAACAGATGCAACAGAAACTATAGATATGCAGC 480
 DB 1521 GGAATACCAACAGAAACTGGACCAATAACAGATGCAACAGAAACTATAGATATGCAGC 1462
 QY 481 ACCTACCTCTCTGAAATAGTGGTTCTTAATCTGTACTTGGGACGGTGTGGAAGT 540
 DB 1461 ACCTACCTCTCTGAAATAGTGGTTCTTAATCTGTACTTGGGACGGTGTGGAAGT 1402
 QY 541 TCAAAAGTTTCAGAAATGAAAGACGTCTCTGTGCTCTCTACTCTTACAAAGAGAACAT 600
 DB 1401 TCAAAAGTTTCAGAAATGAAAGACGTCTCTGTGCTCTCTACTCTTACAAAGAGAACAT 1342
 QY 601 GCTGCCATTTGCCGTGTAGCCAGCTCTCTCTGGAATTTTACACTCGCTGTGTAGATGAT 660
 DB 1341 GCTGCCATTTGCCGTGTAGCCAGCTCTCTCTGGAATTTTACACTCGCTGTGTAGATGAT 1282
 QY 661 GAGCTCTTTGGAGGCCATTTAGCCAAACACCCAGGAGCCAGTTTATGTCTTGTGA 720
 DB 1281 GAGCTCTTTGGAGGCCATTTAGCCAAACACCCAGGAGCCAGTTTATGTGTGTGA 1222
 QY 721 GACACAAGACCAAGATCTGGCATTTCTTGTGCTCATAATGAGAATAGTTCTCCAATTA 780
 DB 1221 GACACAAGACCAAGATCTGGCATTTCTTGTGCTCATAATGAGAATAGTTCTCCAATTA 1162
 QY 781 GCCAAGATGAACCTCATGGACATCAACCAAGATCTTCTCCCTCTCGAGCCCGACAGGAG 840
 DB 1161 GCCAAGATGAACCTCATGGACATCAACCAAGATCTTCTCCCTCTCGAGCCCGACAGGAG 1102
 QY 841 GAGGAGGACATGACACAGAGGAGAGAGGCTCTCAATCAAGCATGTGTATGACACGAC 900
 DB 1101 GAGGAGGACATGACACAGAGGAGAGAGGCTCTCAATCAAGCATGTGTATGACACGAC 1042
 QY 901 TCCTATACCTTTGGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAG 960

Db 1041 TCCTATATCTTGGACACGAGCTTTTGGCCGAGGAGCTTTACAAAGCTTTTCATCAACAGCAGG 982
Qy 961 AGTGGCTGGGGTGTCTTCCCTGGGACACCTTGGCTTGGCTTCTTATGCGGCACATTGAGC 1020
Db 981 AGTGGCTGGGGTGTCTTCCCTGGGACACCTTGGCTTGGCTTCTTATGCGGCACATTGAGC 922
Qy 1021 TGTGTTCAAGTCTCTTATAGCCCATGCTGTGATGTTGACAGCTTGGATGTCGAAGGCACAG 1080
Db 921 TGTGTTCAAGTCTCTTATAGCCCATGCTGTGATGTTGACAGCTTGGATGTCGAAGGCACAG 862
Qy 1081 AGCCCACTTTTCACTGCTGTGATGTCAGTCAGGCGCATCTGGACTGTGTAGCTGTCTTTGGAA 1140
Db 861 AGCCCACTTTTCACTGCTGTGATGTCAGTCAGGCGCATCTGGACTGTGTAGCTGTCTTTGGAA 802
Qy 1141 GCTGGTGCCTCTCTGCTGTGATGTCATCAACAACTGTTCTCCGCTGTCTACAGCTGCC 1200
Db 801 GCTGGTGCCTCTCTGCTGTGATGTCATCAACAACTGTTCTCCGCTGTCTACAGCTGCC 742
Qy 1201 CGTGATGGTGTCTGCTATCTCTGAGGAGCTCTTAGACCATGGTGCAGAGGCCCAACGTC 1260
Db 741 CGTGATGGTGTCTGCTATCTCTGAGGAGCTCTTAGACCATGGTGCAGAGGCCCAACGTC 682
Qy 1261 AAAGCTAAATACCACTCTGGGATCAAAATAGCTTCAATGTTCTGGCCCCCTCTATTGG 1320
Db 681 AAAGCTAAATACCACTCTGGGATCAAAATAGCTTCAATGTTCTGGCCCCCTCTATTGG 622
Qy 1321 GCCGAGCTCTACGGGACCTGGAGCTGTTCCGCTGCTTTTCCGCTGCTTCCGCTGCTGAGCCCT 1380
Db 621 GCCGAGCTCTACGGGACCTGGAGCTGTTCCGCTGCTTTTCCGCTGCTTCCGCTGCTGAGCCCT 562
Qy 1381 GACTCAACTGCACTGACGAGGCGCTTATGGCTGCTGTCGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 561 GACTCAACTGCACTGACGAGGCGCTTATGGCTGCTGTCGCTGCTGCTGCTGCTGCTGCTGCT 502
Qy 1441 GAAATCTGCTCCATCAATATTTGAGCAGAGTATATCAGCTGTTATCGATTTTGGT 1500
Db 501 GAAATCTGCTCCATCAATATTTGAGCAGAGTATATCAGCTGTTATCGATTTTGGT 442
Qy 1501 GCTAATATCTACCTTCCATCTCTCTCCCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 1560
Db 441 GCTAATATCTACCTTCCATCTCTCTCCCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 382
Qy 1561 GCATTGCTGTACAGGCGCGAGCCACTCCAGGCTCACTTCTATCAGAGTTCGTTTGTAGTC 1620
Db 381 GCATTGCTGTACAGGCGCGAGCCACTCCAGGCTCACTTCTATCAGAGTTCGTTTGTAGTC 322
Qy 1621 GTCCGAGAGCTTGTGCGAGGCTGCGCCAGCCCAAGCCATCAACAGCTGGATATTCCT 1680
Db 321 GTCCGAGAGCTTGTGCGAGGCTGCGCCAGCCCAAGCCATCAACAGCTGGATATTCCT 262
Qy 1681 CCCATCTGATTTAGCTACCTTAAACACCACTGTTATCTGCGAGTCTCCCGAGGACTTA 1740
Db 261 CCCATCTGATTTAGCTACCTTAAACACCACTGTTATCTGCGAGTCTCCCGAGGACTTA 202
Qy 1741 TGATGCTCTCGAAACACCTCTGGGAGCTCACGTTGAGAGCATTCACAGCTCATCCA 1800
Db 201 TGATGCTCTCGAAACACCTCTGGGAGCTCACGTTGAGAGCATTCACAGCTCATCCA 142
Qy 1801 CTTACTGGAGCTGCTCTCTGTTATTATCTTCCACAAATAAAATTTCTCCAGAAATAAGTA 1860
Db 141 CTTACTGGAGCTGCTCTCTGTTATTATCTTCCACAAATAAAATTTCTCCAGAAATAAGTA 82
Qy 1861 AAAAAAAAAAAAAAAAAA 1878
Db 81 AAAAAAAAAAAAAAAAAA 64

RESULT 5
ABK1107
ID ABK11107 standard; DNA; 1222 BP.
XX
AC ABK11107;
XX

DT 05-JUN-2002 (first entry)
XX DNA encoding human NOV5b protein, homologue of ankryrin-repeat proteins.
DE Human; NOVX-associated disorder; developmental disorder; blood disorder;
XX endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
KW respiratory disorder; inflammatory disorder; reproductive disorder;
KW neurodegenerative disorder; autoimmune disorder; infectious disease;
KW cardiovascular disorder; cell signal processing; ankryrin-repeat protein;
KW metabolic pathway modulation; NOV5b; gene; ds.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT 6..1088
CDS /*tag= a
FT /*product= "NOV5b"
FT
XX WO200206329-A2.
PN 24-JAN-2002.
PD
XX 18-JUL-2001; 2001WO-US22709.
PF 18-JUL-2000; 2000US-218870P.
XX 18-JUL-2000; 2000US-218875P.
PR 18-JUL-2000; 2000US-218901P.
PR 24-JUL-2000; 2000US-220273P.
PR 26-JUL-2000; 2000US-220912P.
PR 27-JUL-2000; 2000US-221233P.
PR 28-JUL-2000; 2000US-221650P.
XX (CURA-) CUPAGEN CORP.
XX
XX Rastelli L, Shinkets RA, Zerhusen B, Malyankar UM, Padigaru M;
PI WPI; 2002-179781/23.
XX P-PSDB; AAU77411.
DR
XX Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT polypeptides, useful for treating developmental disorders, endocrine
PT disorders, vascular disorders, infectious diseases and
PT neurodegenerative disorders -
XX
PS Claim 9; Page 39; 178pp; English.
XX
CC The present invention relates to the isolation of novel human
CC polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a,
CC NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC The NOVX polypeptides are related to NOPE, cadherin, interferon
CC alpha-13, ADAM, ankryrin repeat-containing, transpanin or semaphorin
CC polypeptides. The sequences of the invention are useful for identifying
CC an agent (a cellular receptor or downstream effector) that binds to
CC the NOVX polypeptide, or an agent that modulates it's expression or
CC activity. They are useful for treating or preventing NOVX-associated
CC disorders such as developmental disorders, endocrine disorders, vascular
CC diseases, gastrointestinal disorders, reproductive disorders, inflammatory
CC disorders, blood disorders, autoimmune and immune disorders, infectious diseases,
CC cardiovascular disorders, cancers, and other disorders related to cell
CC signal processing and metabolic pathway modulation. The present sequence
CC encodes the human NOV5b protein.
XX
SQ Sequence 1222 BP; 276 A; 352 C; 278 G; 316 T; 0 other;

Query Match 59.5%; Score 1117.8; DB 24; Length 1222;
Best Local Similarity 99.8%; Pred. No. 2.2e-306;
Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 730 CCAAAGATCTGCATTCTCTGCTCATATGAGAATAGTTCTCAATAGCCAGATG 789
Db 102 CCACAGATCTGCATTCTCTGCTCATATGAGAATAGTTCTCTCAATAGCCAGATG 161

```
QY 790 AACCTCATGACATCACCAGATCTTCTCCTCTCGAGCCCGACAGGAGGAGGAC 849
Db 162 AACCTCATGACATCACCAGATCTTCTCCTCTCGAGCCCGACAGGAGGAGGAC 221
QY 850 ACTGACACAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACACGACTCTCTACT 909
Db 222 ACTGACACAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACACGACTCTCTACT 281
QY 910 TTGGACAGCTTTTGGCCAGGAGCGTTACAAAGCTTTTCAACAGCAGGAGTGTGCTGG 969
Db 282 TTGGACAGCTTTTGGCCAGGAGCGTTACAAAGCTTTTCAACAGCAGGAGTGTGCTGG 341
QY 970 GGTGTTCTGGGACACCTTGGCTGGCTCTTCTATGGCCACTTGAGCTTTTGCAA 1029
Db 342 GGTGTTCTGGGACACCTTGGCTGGCTCTTCTATGGCCACTTGAGCTTTTGCAA 401
QY 1030 GTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGGCCACTT 1089
Db 402 GTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGGCCACTT 461
QY 1090 TTCACTGCTGTAGTCAATGAGCCATCTGGAGCTGTAGCTGTCTTTTGAAGCTGTGCTC 1149
Db 462 TTCACTGCTGTAGTCAATGAGCCATCTGGAGCTGTAGCTGTCTTTTGAAGCTGTGCTC 521
QY 1150 TCTCTGGTGTAGCTCTACAAACACTGTTCTCCGCTGTCTACAGCTGCCCTGTAGTGT 1209
Db 522 TCTCTGGTGTAGCTCTACAAACACTGTTCTCCGCTGTCTACAGCTGCCCTGTAGTGT 581
QY 1210 GCTGTTGCTATCTGACGAGCTCTTAGACCATGGTGCAGAGCCCAACGTCAAAGCTAA 1269
Db 582 GCTGTTGCTATCTGACGAGCTCTTAGACCATGGTGCAGAGCCCAACGTCAAAGCTAA 641
QY 1270 CTACCACTCTGGGCATCAACATAGCTTCATGTTCTGGCCCTCTATTGTCGCGAGTC 1329
Db 642 CTACCACTCTGGGCATCAACATAGCTTCATGTTCTGGCCCTCTATTGTCGCGAGTC 701
QY 1330 TACGGGCACCTGGAGCTGTTTCCGCTGCTTTTCTCCAGCGGCGAGACCTCTGACTACAC 1389
Db 702 TACGGGCACCTGGAGCTGTTTCCGCTGCTTTTCTCCAGCGGCGAGACCTCTGACTACAC 761
QY 1390 TGCACTGACAGGCGCTATTGGCTGTGTCGTCGACGACCCGACCTCTCTTGAATCTGC 1449
Db 762 TGCACTGACAGGCGCTATTGGCTGTGTCGTCGACGACCCGACCTCTCTTGAATCTGC 821
QY 1450 CTCATCATATGTCGACGAGTATATCCAGCTGTATGATGTTTGGTGTATATC 1509
Db 822 CTCATCATATGTCGACGAGTATATCCAGCTGTATGATGTTTGGTGTATATC 881
QY 1510 TACCTTCCATCTCTCTCCCTTGAACCTGACCTCAAGATGATAAAGGCATTGCTGCTG 1569
Db 882 TACCTTCCATCTCTCTCCCTTGAACCTGACCTCAAGATGATAAAGGCATTGCTGCTG 941
QY 1570 CTACAGGCGGAGCCACTCAAGCTCACTTCTATCAAGCTCGTTCGTTAGTGTGCGCAGA 1629
Db 942 CTACAGGCGGAGCCACTCAAGCTCACTTCTATCAAGCTCGTTCGTTAGTGTGCGCAGA 1001
QY 1630 GCCTTGTGCGAGCTGCGCAGCCACAGCCATCAACAGCTGATATCTCCCATGTTG 1689
Db 1002 GCCTTGTGCGAGCTGCGCAGCCACAGCCATCAACAGCTGATATCTCCCATGTTG 1061
QY 1690 ATTAGCTACCTAAACACCACTGATATCTTGCAGTCTCCAGGAACTTATGATCCCTC 1749
Db 1062 ATTAGCTACCTAAACACCACTGATATCTTGCAGTCTCCAGGAACTTATGATCCCTC 1121
QY 1750 CGAAACACCTGGGAGCTCACTGAGCTGGAGAGCATTCAGCTCATCCACTTACCTGG 1809
Db 1122 CGAAACACCTGGGAGCTCACTGAGCTGGAGAGCATTCAGCTCATCCACTTACCTGG 1181
QY 1810 AGCTGCTCTCTGATATCTTATCTCCCAATAAATCTCCAG 1850
Db 1182 AGCTGCTCTCTGATATCTTATCTCCCAATAAATCTCCAG 1222
```

```
RESULT 6
ID ABZ11533 standard; cDNA; 1127 BP.
XX ABZ11533;
AC ABZ11533;
XX 20-JAN-2003 (first entry)
DT Human polynucleotide SEQ ID NO 415.
DE
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
cell-proliferative disorder; neurodegenerative disease; bacterial;
Parkinson's disease; Alzheimer's disease; autoimmune disease;
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;
antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200270539-A2.
XX 12-SEP-2002.
XX 05-MAR-2002; 2002WO-US05095.
XX 05-MAR-2001; 2001US-0799451.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;
Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
DR P-PSDB; ABP69316.
XX
PT New polynucleotides comprising sequences assembled from expressed
sequence tags (ESTs), useful for treating cell-proliferative,
neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
platelet or coagulation disorders
XX
PS Claim 1: SEQ ID NO 415; 1012pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (I) comprising a
nucleotide sequence selected from any of 948 sequences
(ABZ1119-ABZ12066) or their mature protein coding portion, active domain
coding protein or complementary sequences. The polynucleotides are useful
for identifying expressed genes or for physical mapping of human genome.
The encoded polypeptides (ABP6902-ABP6949) are useful as molecular
weight markers, as a food supplement, for generating antibodies, in
medical imaging, screening and diagnostic assays and for treating
cell-proliferative disorders (cancer), neurodegenerative diseases
(Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
disorders, platelet or coagulation disorders, wound, burns, incision,
ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
parasitic), arthritis, etc.
CC
CC Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1127 BP; 244 A; 329 C; 281 G; 273 T; 0 other;
Query Match 52.7%; Score 989.8; DB 24; Length 1127;
Best Local Similarity 99.8%; Pred. No. 4.5e-270;
Matches 991; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 735 GATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCATTAGCCAAATGAACT 794
Db 134 GATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCAAATGAAATGAACT 193
```


XX		Phosphatase; gene therapy; enzyme; cancer; pathophysiological hypoxia;
KW		cardiac dysfunction; vascular disorder; myopathy; ectodermal dysplasia;
KW		congenital muscle disorder; Papillon-Lefevre syndrome; Cowden disease;
KW		Moebius syndrome; Bjornstad syndrome; Bannayan Zonana syndrome;
KW		schizophrenia; hamartoma; ss.
OS		Homo sapiens.
XX		
PN		WO200112819-A2.
PD		22-FEB-2001.
PF		11-AUG-2000; 2000WO-US222158.
PR		13-AUG-1999; 99US-0149005.
PR		(SUGE-) SUGEN INC.
PA		
XX		
PI		Plowman GD, Martinez R, Whyte D, Hill R, Flanagan P, Lioubin M;
DR		WPI: 2001-211226/21.
XX		P-PSDB; AAB73229.
DR		
XX		
PT		New protein phosphatase polypeptide for diagnosing and treating
PT		phosphatase related disorders such as cancer, schizophrenia, cardiac
PT		dysfunction and/or vascular disorders -
XX		
PS		Claim 1; Fig 4; 138pp; English.
XX		
CC		The present invention relates to phosphatase proteins and coding
CC		sequences. The present sequence is one such phosphatase coding sequence.
CC		Phosphatases are enzymes that catalyse the dephosphorylation of proteins
CC		modified by phosphorylation of serine, threonine or tyrosine residues.
CC		The phosphatases are useful for treating a variety of diseases: for
CC		example cancer e.g. breast, urogenital, prostate, head, neck, lung
CC		cancers, synovial sarcomas, renal cell carcinoma, non-small cell lung
CC		cancer, hepatocellular carcinoma, pancreatic endocrine tumours, stomach
CC		cancer, glioblastoma, colorectal cancer and thyroid cancer.
CC		pathophysiological hypoxia, cardiac dysfunction and/or vascular
CC		disorders, myopathies, congenital muscle disorders, Papillon-Lefevre
CC		syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome,
CC		Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and
CC		hamartomas.
XX		
SQ		Sequence 2353 BP; 670 A; 506 C; 535 G; 642 T; 0 other;
		Query Match 12.9%; Score 241.8; DB 22; Length 2353;
		Best Local Similarity 58.3%; Pred. No. 1.2e-57;
		Matches 423; Conservative 0; Mismatches 302; Indels 0; Gaps 0;
QY	10	GCGACGTTGGCACAGGGGTAGAACAACCTGAATGGTGGTCGTTATGTGAGTAGAATA 69
Db	239	GCTCAGGGAAGCCGAAAGGTTGAAAATGTCGCTTGTTAGTCGAGTGTCTCTCAAAAAA 298
QY	70	CCAGCTTAATGGCATCTTTATTCTACTGCCAACCCACCTGATCTATGTGGAGGCTTCAGGT 129
Db	299	GCAGCTCTAGGTACTTTGTTATTTGACGCTACCACATGTCATATTCGTGGAAAAATTCACCT 358
QY	130	GCAGCCCGGAAGAAACAATGGATTGCATCCATCACAATTGCCACTGTGGAGAAAGTTACCC 189
Db	359	GACGCAAGAAAAAGAAACAATGGATTCTTCACAGTCAGATTTCCACCATTGAGAAACAGGCA 418
QY	190	ATCACTAGCTGGTGTGTCCTCGCTGACCTCGCTGCAAGAAATTCGGGTGGCCGCACTTT 249
Db	419	ACAAACGCTACCGGATGCCCTCTGCTGATTCGCTGCAAGAACTTTTCAGATAATACAGCTC 478
QY	250	GTTTTAGATTCTGACCTTGTGTGTCATCAGGTTTATATTTCACTGCTCAAGCTTTCTCAG 309
Db	479	ATCATACCTCAGAAAGAGATTGCCACGAGTGATCATCTCCCTGATACGCTTTGCAAGG 538
QY	310	CCAGCATTTACCTGAAGATCTTTATGCTTTTTTCTTTTATTAATCCCAAATCCTCAAAGAGATG 369

Db 634 GACGCTGTGCTCCATGAACGAGAGTACGAGCTGTGCGACACCTATCCGCGTCAGATTTA 693
QY 501 GGTTCCTAAATCTGTTACCTTGGGAACGGTGGTTGGAAGTTCAAAGTTTCAAGATGAAGA 560
Db 694 CGTGCCCAAGAGGCCACACGCTGATGCTCATCAGCAGCTCGGATTCGCTCCAGGG 753
QY 561 ACGTGTCCTGTGCTCTCTACCTCTTCAAAAGAGAACAAATGCTGCCATTTGCCGCTGTAG 620
Db 754 GCGGCTGCCAGTGTCTACCTATCTGCACAAC---AACAGGCTTCATCTGCCGCTGCAG 810
QY 621 CGAGCTCTCTCTGGATTTTACACTCGCTGTGTAGATGAGTCTTTGTTGGAGCCAT 680
Db 811 TCAGCCCTCTCCGGATTCTAGTCCCGCTCTCTGGAGGATGAGCAGATGCTGGAGGCCAT 870
QY 681 TAGCAAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTAGACACACCAAGAT 737
Db 871 ACGCAGACGANTCCACAGGACTACATGATGTGTGGTGACACGCGACCCGGAT 927

RESULT 13
ID ABA19208/c
XX ABA19208 standard; DNA; 10828 BP.
AC ABA19208;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 11539.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antileumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antituberc; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-541565/60.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure; SEQ ID NO 11539; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
 CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 10828 BP; 3640 A; 2147 C; 1989 G; 3052 T; 0 other;
 Query Match 9.0%; Score 169; DB 22; Length 10828;
 Best Local Similarity 92.2%; Pred. No. 1.2e-36;
 Matches 178; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 282 TTATATTTTCACTGCTCAAGCTTTTCTCAGCCAGCATTACCTGAAGATCTTTATGCTTTTTC 341
 |||||
 DB 6241 TTTTGTGTTTCCATGCTATCTCTCTCTCCAGCATTACCTGAAGATCTTTATGCTTTTTC 6182
 |||||
 QY 342 TTATATTTTCCAAATCTCTCAAAAGAGATGAGGAAAGTGGATGGAAGTGAATGACCCCAAT 401
 |||||
 DB 6181 TTATATTTTCCAAATCTCTCAAAAGAGATGAGGAAAGTGGATGGAAGTGAATGACCCCAAT 6122
 |||||
 QY 402 ATCAGACTTTTGGCGTATGCGGAATACCCAAAGAGAACTGGACCATAACAGATGCCAACAG 461
 |||||
 DB 6121 ATCAGACTTTTGGCGTATGCGGAATACCCAAAGAGAACTGGACCATAACAGATGCCAACAG 6062
 |||||
 QY 462 AAACATATGAGATA 474
 |||||
 DB 6061 AAACATATGAGGTA 6049
 |||||
 RESULT 14
 AAK70743/c
 ID AAK70743 standard; DNA; 10828 BP.
 XX
 AC AAK70743;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25555.
 XX
 DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 KW
 XX Homo sapiens.
 OS
 XX
 XX WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0215647.
 PR 07-JUL-2000; 2000US-0215880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226861.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.

```
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 2555; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM92170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 10828 BP; 3640 A; 2147 C; 1989 G; 3052 T; 0 other;
Query Match
Best Local Similarity 9.0%; Score 169; DB 22; Length 10828;
Matches 178; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 282 TTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTACTGAGATCTTTATGCTTTTTC 341
Db 6241 TTTTGTTCATGCTATCTCTCTCTCCAGCATTACTGAGATCTTTATGCTTTTTC 6182
QY 342 TTATAATCCCAATCTCTCAAAAGAGATGAGGAAAGTGGATGGAACCTGATGACCCAAAT 401
Db 6181 TTATAATCCCAATCTCTCAAAAGAGATGAGGAAAGTGGATGGAACCTGATGACCCAAAT 6122
```

QY 402 ATCAGACTTTGGCGGTATGGGATATCCCAACAGAAACTGGACCATACAGATGCCAACAG 461
 DB 6121 ATCAGACTTTGGCGGTATGGGATATCCCAACAGAAACTGGACCATACAGATGCCAACAG 6062
 QY 462 AACTATGAGATA 474
 DB 6061 AACTATGAGGTA 6049

RESULT 15

ABL17297
 ID ABL17297 standard; DNA; 3169 BP.

AC ABL17297;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 3364.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Claim 1; SEQ ID NO 3364; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3169 BP; 816 A; 877 C; 796 G; 680 T; 0 other;

Query Match 8.0%; Score 149.6; DB 23; Length 3169;
 Best Local Similarity 54.6%; Pred. No. 2e-31;
 Matches 321; Conservative 0; Mismatches 264; Indels 3; Gaps 1;

QY 150 GATTGACATCCATCATTGCGCTGATGGAGAGTTACCCATCATTAGCTGGTTGTCC 209

DB 726 GATCTGCGACATGCGCTGCGGCTGAGAGCTTCCCTTGAGCAGCAGGATCTCC 785

QY 210 CTGACCCCTCGGTGCAAGATTTCGGGTGGCCACTTGTGTTAGATTCTGACCTTGT 269

DB 786 GCTACTCATCCGCTGCAAGACCTTCTCTCCGTAATTTTCGTCATTCCCAAGACTCCGA 845

QY 270 GTGCCATGAGGTTTATATTCTACTGCTCAAGCTTTCTCAGCCAGCATTTACCTGAAGATCT 329

DB 846 GTGCCACGATGCTACACTTCGCTGCTGAAACTCTTCCAGCCGGTGTCCATCAACAAATT 905
 QY 330 TTATGCTTTTCTTATAATCCCAATCTCRAAAGAGATGAGGGAAGTGGATGGAACT 389
 DB 906 GTACTGCTTCAACTACCGCCGAACAAGGATGATTTCCCAAGAAGCGCTGGCTGGGATTA 965
 QY 390 GATTGACCCCAATPATCAGACTTTTGGCGGTATGGGAATACCCAAACAGAAACTGGACCATTAAC 449
 DB 966 CTTTAAACTGGAGGCGGAGTTCAAGCACATGCTGGTGGCCCAACGAGGCCCTGGACGCTGTG 1025
 QY 450 AGATGCCAACAGAAACTATGAGATATGACAGACCTACCTCTCTGAAATAGTGTTCCTAA 509
 DB 1026 CTCCATGAACGAGAAAGTAGAGCTGTGCGACACCTATCCGCGTTCAGATTTACGTGCCCAA 1085
 QY 510 ATCTGTTACCTTTGGGAACGGTGGTGGAGTTTCAAAAGTTTCAGAAAGTAAAGAACGTGTCCC 569
 DB 1086 GGAGGCCACCGCTGATGCTCATCAGCAGCTCGGATTCGCTCCAAGGGGGGCTGCC 1145
 QY 570 TGTGCTCTCTACCTCTTACAAAGAGAACAATGCTGCCATTTTCGCGCTGTAGCCAGCTCT 629
 DB 1146 AGTGTCTACCTATCTGCACAAC--AACAAAGCTTCCATCTGCGCTGCAGTCAGCCCT 1202
 QY 630 CTCTGATTTTACACTCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTAGCCAAAC 689
 DB 1203 GTCCGGATTCAGTGGCCCGCTGTCTGGAGGATGAGCAGATGCTGGAGGCCATTAGCAAGAC 1262
 QY 690 AAACCCAGGAGCCAGTTTATGCTATGTTGTACACACACAGACCAAGAT 737
 DB 1263 GAATTCCAACCGGACTACATGTATGTGTGACACACGCGCCGGAT 1310

Search completed: December 1, 2003, 10:38:15

Job time : 516 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:53 ; Search time 39 Seconds
(without alignments)
541.410 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLDPIDFGRMG.....QAINQLDIPPLMLISYLKQL 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	68.2	308	1 AS12 HUMAN	Q8wxk4 homo sapien
2	499	21.3	335	1 ASB1 HUMAN	Q9y576 mus musculus
3	481	20.5	336	1 ASB1 MOUSE	Q9wv74 mus musculus
4	296.5	12.7	643	1 MTR2 HUMAN	Q13614 homo sapien
5	288	12.3	391	1 MTR7 HUMAN	Q9y216 homo sapien
6	287	12.2	465	1 MTR6 HUMAN	Q9y217 homo sapien
7	283	12.1	669	1 MTR1 MOUSE	Q9z2c4 mus musculus
8	281	12.0	662	1 MTR1 HUMAN	Q13613 homo sapien
9	270.5	11.5	278	1 AS13 HUMAN	Q8wxk3 homo sapien
10	264	11.3	603	1 MTM1 HUMAN	Q13496 homo sapien
11	261.5	11.2	329	1 ASB5 HUMAN	Q8wxk0 homo sapien
12	258	11.0	603	1 MTM1 MOUSE	Q9z2c5 mus musculus
13	250	10.7	290	1 ASB9 MOUSE	Q91z78 mus musculus
14	246	10.5	323	1 AS11 HUMAN	Q8wxh4 homo sapien
15	243.5	10.4	518	1 ASB3 HUMAN	Q9y575 homo sapien
16	241	10.3	294	1 ASB3 HUMAN	Q96dx5 homo sapien
17	228.5	9.8	525	1 ASB3 MOUSE	Q9wv72 mus musculus
18	224.5	9.6	1862	1 ANK1 MOUSE	Q02357 mus musculus
19	224.5	9.6	1880	1 ANK1 HUMAN	P16157 homo sapien
20	215	9.2	1166	1 TNK2 HUMAN	Q9h2k2 homo sapien
21	210.5	9.0	688	1 YJ80 YEAST	P47147 saccharomyc
22	208.5	8.9	1327	1 TNK1 HUMAN	Q95271 homo sapien
23	205	8.7	4377	1 ANK3 HUMAN	Q12555 homo sapien
24	202.5	8.6	3924	1 ANK2 HUMAN	Q01484 homo sapien
25	202	8.6	768	1 YB23 HUMAN	Q9ulj7 homo sapien
26	199.5	8.5	288	1 ASB8 MOUSE	Q91zt9 mus musculus
27	199.5	8.5	587	1 ASB2 HUMAN	Q96q27 homo sapien
28	194	8.3	583	1 AS15 MOUSE	Q8vhs6 mus musculus
29	193.5	8.3	429	1 AS10 HUMAN	Q8wx13 homo sapien
30	190.5	8.1	1059	1 Y379 HUMAN	O15084 homo sapien
31	188.5	8.0	288	1 ASB6 HUMAN	Q9h765 homo sapien
32	187	8.0	410	1 V240 FOWPV	P14360 fowlpox vir
33	184.5	7.9	434	1 AS15 HUMAN	Q8wxk1 homo sapien

34	184.5	7.9	434	1 V023 FOWPV	Q915sh fowlpox vir
35	179	7.6	434	1 V219 FOWPV	Q91516 fowlpox vir
36	177.5	7.6	592	1 V246 FOWPV	Q91424 fowlpox vir
37	177.5	7.6	1134	1 ANS1_HUMAN	Q92625 homo sapien
38	177.5	7.6	1150	1 ANS1_MOUSE	P59672 mus musculus
39	172.5	7.4	318	1 ASB7_MOUSE	Q91zu0 mus musculus
40	170	7.3	426	1 ASB4_HUMAN	Q9y574 homo sapien
41	169	7.2	426	1 ASB4_MOUSE	Q9wv71 mus musculus
42	168.5	7.2	735	1 RNSA_MOUSE	Q05921 mus musculus
43	168	7.2	528	1 PSCA_HUMAN	Q96134 homo sapien
44	167	7.1	323	1 ANKH_CHRVI	Q08527 chromatinum
45	167	7.1	453	1 AS16_HUMAN	Q96ns5 homo sapien

ALIGNMENTS

RESULT 1
AS12_HUMAN
ID AS12_HUMAN STANDARD; PRT; 308 AA.
AC Q8wxk4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Ankyrin repeat and SOCS box containing protein 12 (ASB-12).
GN ASB12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kile B.T., Nicola N.A.;
RT "SOCS box proteins."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -!- SIMILARITY: Contains 1 SOCS box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF403030; AAL57349.1; -
CC Genew; HGNC:19763; ASB12.
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; ank; 5.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 5.
CC PROSITE; PS50088; ANK_REPEAT; 4.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC KW ANK repeat; Repeat.
CC FT REPEAT 63 92 ANK 1.
CC FT REPEAT 96 125 ANK 2.
CC FT REPEAT 129 158 ANK 3.
CC FT REPEAT 171 200 ANK 4.
CC FT REPEAT 213 243 ANK 5.
CC SQ SEQUENCE 308 AA; 33829 MW; 80C47C544D03D8AC CRC64;

Query Match 68.2%; Score 1599; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 3e-118;
Matches 308; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	141	MNLMDITKIFSLQDPKKEEDTDEKQALNAVYDNDSTYLDQLLRQRYKRFNSRSG	200
Db	1	MNLMDITKIFSLQDPKKEEDTDEKQALNAVYDNDSTYLDQLLRQRYKRFNSRSG	60
Qy	201	MGVPGTPLRLAASGYHLSCLQVLLAHGADVDSLDVKAQTPFLTFVSHGHLCDCVRVLEAG	260
Db	61	MGVPGTPLRLAASGYHLSCLQVLLAHGADVDSLDVKAQTPFLTFVSHGHLCDCVRVLEAG	120

QY 261 ASPGSGIYNNCSVLTAAADGAVAILQELLDHGAENAVKAKLPVWASNTASCSGPLYLAA 320
 Db 121 ASPGSGIYNNCSVLTAAADGAVAILQELLDHGAENAVKAKLPVWASNTASCSGPLYLAA 180
 QY 321 VVGHLCDFRLLHGHADPDYNCDDQGLLARVPRTLLLEICLHNCPEYIQLLIDFGAN 380
 Db 181 VVGHLCDFRLLHGHADPDYNCDDQGLLARVPRTLLLEICLHNCPEYIQLLIDFGAN 240
 QY 381 IYLPISLSDLTSDODDGIALLQARATPRSLLSQVRLVVRALCQAGQPOAINQLDIPM 440
 Db 241 IYLPISLSDLTSDODDGIALLQARATPRSLLSQVRLVVRALCQAGQPOAINQLDIPM 300
 QY 441 LISYLKHQ 448
 Db 301 LISYLKHQ 308

RESULT 2
 ASB1_HUMAN STANDARD; PRT; 335 AA.
 AC Q9Y576; Q9ULS4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 1 (ASB-1).
 GN ASB1 OR KIAA1146.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20564172; PubMed=11111040;
 RA Kile B.T., Viney E.M., Willson T.A., Brodnicki T.C., Cancilla M.R.,
 RA Herlihy A.S., Croker B.A., Baca M., Nicola N.A., Hilton D.J.,
 RA Alexander W.S.;
 RT "Cloning and characterization of the genes encoding the ankyrin repeat
 RT and SOCS box-containing proteins Asb-1, Asb-2, Asb-3 and Asb-4.";
 RL Gene 258:31-41(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulláhy S.J.,
 RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Nadeau A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 65-335 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039618; PubMed=10574461;
 RA Hiroseawa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
 RA Ohara O.;
 RT "Characterization of cDNA clones selected by the GeneMark analysis
 RT from size-fractionated cDNA libraries from human brain.";
 RL DNA Res. 6:329-336(1999).

CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -!- SIMILARITY: Contains 1 SOCS box domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF156777; AAD41894.1; -;
 DR EMBL; BC014528; AAH14528.1; -;
 DR EMBL; AB032972; BAA86460.1; -;
 DR MIM; HGNC:16011; ASB1.
 DR MIM; 605758; -;
 DR HSSP; Q00420; IAWC.
 DR GO; GO:0005622; C:intracellular; IC.
 DR GO; GO:0042036; P:negative regulation of cytokine biosynthesis; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001496; SOCS.
 DR Pfam; PF00023; ank; 6.
 DR SMART; SM00248; ANK; 5.
 DR PROSITE; PS50088; ANK REPEAT; 3.
 DR PROSITE; PS50297; ANK REP REGION; 1.
 DR PROSITE; PS50225; SOCS; 1.
 KW ANK repeat; Repeat.
 FT REPEAT 36 68 ANK 1.
 FT REPEAT 77 106 ANK 2.
 FT REPEAT 110 139 ANK 3.
 FT REPEAT 143 172 ANK 4.
 FT REPEAT 191 220 ANK 5.
 FT REPEAT 235 265 ANK 6.
 FT DOMAIN 286 335 SOCS BOX.
 SQ SEQUENCE 335 AA; 37014 MW; 0843C96AD1AF60D9 CRC64;
 Query Match 21.3%; Score 499; DB 1; Length 335;
 Best Local Similarity 39.1%; Pred. No. 6.9e-32;
 Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;
 QY 166 EKQALNQAVYDNDSTYLDQLLRQERYKRFINSRSGW---GVFGTPLRLAASYGHLSCLOV 222
 Db 37 EDTRLHDAAYVGLDQLTSLRLQEEYSRINEKSVGCCWLPCTPLRIAATAAGHSCVDF 96
 QY 223 LIAHGADVDSLDVKAQTPLFTAVSHGHLDCTVRLLEAGASPGSGIYNNCSVLTAAADGA 282
 Db 97 LIRKGAEDLVGVKQGTALYAVAVNGHLESTQILLEAGADPNGRHRSRTPVYHASRVGR 156
 QY 283 VAILQELLDHGAENAVKAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLH 334
 Db 157 ADILKALIRYGDVVDVNNHLLTDPVQFRFSRLTSLVVC--PLYISAAVHNLCQFRLLH 214
 QY 335 GADPDYNC---TDQGLLARVPRTLLLEICLHNCPEYIQLLIDFGANIYL---PSLS 387
 Db 215 GANPDYNCNPVNTQGYR--GSPGCVMDAVLRHGCFAAFVSLVFEFGANLVKWSLIG 272
 QY 388 LDLTQ---DDKGIALLQARATPRSLLSQVRLVVRALCQAGQPOAINQLDIPMLISY 444
 Db 273 PESRGRKRVDPALQVFKEARSPVRTLCLCRVAVRRALGK-HRLHLPLSLPLPPIKPF 331
 QY 445 LKHQ 448
 Db 332 LUHE 335

RESULT 3
 ASB1_MOUSE
 ID ASB1_MOUSE STANDARD; PRT; 336 AA.
 AC Q9WV74;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 1 (ASB-1).


```
Query Match 12.7%; Score 296.5; DB 1; Length 643;
Best Local Similarity 25.4%; Pred. No. 1.3e-15; Indels 119; Gaps 17;
Matches 100; Conservative 57; Mismatches 118;

QY 3 ESKWKLDIPISFGRMGIPNRNWTITDANRNYEICSITYPEIIVPKSVTLGTIVGSSKFR 62
   :|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 ENGKLYDPLLEVRROGIPNESWRITKINERYELCDTYPALLVVPANIPDEELKRVASR 262
   :|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 63 SKERVPLSVLYKNNAAICRCSQPLSGPY-TRCVDELLLEAISQTNPGSOFMYVVDPR 121
   :|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 SRGRIPLSWIHPESQATITRCSQPMVGSGRKSDEKYLQAIMDSNAQSHKIFIFDAR 322
   :|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 122 PKI-----WHFLVLMIRVLQALQNM--DITKIFSLLOPDKEE--- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 PSVANAVAKGGGYESEDAYQNAELVF---LDIHNHVRMSLRKLKEIVYPNIEETHW 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 160 ----EDT-----DTEKQALNQAAYDNDV-----TLDQLLRQERYK 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 LSNLESTHLEHIKILLAGALRIADKVESGKTSVVVHCSDGWDRTAQLTSLAMLMLDGY 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 193 RFINS-----RSGWGVGPTPLRAASYHGLSCLQVLLAHG-----ADVSLDVKAQPLFT 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 RTIRGEVLVEKEW-----LSFGHR--FQLRVGHHGDKNHADAD-----RSPVEL 481
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 244 AVSHGHLCVRLVLEAGASPGGSIVNCSFVLTAARDGAVAILQELLDH----- 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 Q-----FIDCVQWQNT--QFTAFEFNE-----YFLITILDHLVSLCLFGFTFLC 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 293 -----GAENYAKKPLVWNA-----SNIASCSGPLY 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 NSEQOQGENLPKRTVSLASYNLSQLEDFTNPLY 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
MTR7 HUMAN
ID MTR7 HUMAN STANDARD; PRT; 391 AA.
AC Q9Y2L6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myotubularin related protein 7 (EC 3.1.3.48) (Fragment).
GN MTR7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,
RA Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
RT gene family from yeast to human.";
RL Hum. Mol. Genet. 7:1703-1712(1998).
CC -!- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
GN MTR7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,
RA Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
RT gene family from yeast to human.";
RL Hum. Mol. Genet. 7:1703-1712(1998).
CC -!- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -!- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF073482; AAC77820.1; -.
DR Genew; HGNC:7454; MTR7.
DR MIM; 603562; -.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.

Query Match 12.2%; Score 287; DB 1; Length 465;
Best Local Similarity 61.6%; Pred. No. 4.6e-15;
Matches 53; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

QY 39 TYPPEIVPKSVTLGTIVGSSKFRKERVPLSVLYKNNAAICRCSQPLSGFYTRCVDD 98
```

Db 2 TYPELVPIASPIITVSSKGRFPVLSYVHODKEAAICRCSQPLSGFSARCLD 61
Qy 99 ELLEALISQTPGSOQFMVVDTRPKI 124
Db 62 EHLQALISKANPVNRYMYVMDTRPKL 87

RESULT 7

MTRL_MOUSE
ID MTRL_MOUSE STANDARD; PRT; 669 AA.
AC Q922C4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myotubularin-related protein 1.
GN MTMRI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buji-Bello A., Tentler D., Kretz C., Dahl N.,
Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
gene family from yeast to human";
RL Hum. Mol. Genet. 7:1703-1712(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Wiehe T., Zhao W., Herman G.E., Rosenthal A., Platzter M.;
RT "Comparative sequence analysis of the mouse Mtm locus and the
corresponding region of human Xq28";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -1- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF073997; AAC77822.1; -
DR EMBL; AF125314; AAF22122.1; -
DR MGD; MGI:1858271; Mtmr1.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0004438; F:phosphatidylinositol-3-phosphatase activity; IDA.
DR InterPro; IPR004182; GRAM dom.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF02893; GRAM; 1.
DR SMART; SM00568; GRAM; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
KW Hydrolase.
FT ACT SITE 442 442 BY SIMILARITY.
SQ SEQUENCE 669 AA; 75313 MW; 1856792245F2D800 CRC64;

Query Match 12.18; Score 283; DB 1; Length 669;
Best Local Similarity 27.5%; Pred. No. 1.6e-14;
Matches 86; Conservative 42; Mismatches 93; Indels 92; Gaps 12;
Qy 4 SGWKLIDIPDGRMGIPNRTWTTIDANRNVICSTYPPPIVVPKSTLTGTVVGSSKFRS 63
Db 229 NGWKVDPVSEYKQGFNENSWKISKINSNVEFCFTPAIVFTSVKDDDLKVAAPRA 288
Qy 64 KERPVLISLYKENNAACRCSQPLSG - FYTRCVDDLELLLEAISQTPGSOQFMVVDTR - 121
Db 289 KGRVPVLISWHIPESQATITRCSQPLVGPNDRCKEDEKYLQTIMDANAQSHKLTFIDRQ 348

Qy 122 -----PKIWHFLVLMIRVLQAKNLM--DITKIFSLQPKDEE 159
Db 349 NSVADTNKAKGGYENESAYPN-----AELIFLEIHNIHVRESLRKLKEIVYPSIDB 401
Qy 160 ED-----TDTTEKQALNQAQVYDNDY-----TLQQLLRQRE 189
Db 402 SHWLSNVDGTHWLEIVIRVLLAGAVRIADKIESGKTSVVIHCSGDWDRTSQLTSLAMLMD 461
Qy 190 RYKRFINS-----RSGWGVPGTFLRLAASYGHLSCLQVLLAHG-----ADVDSLQVKAQTP 240
Db 462 SYRTIKNGFEALIEKEW-----ISFGHREFALRV--GHGDDNHADAD-----RSP 503
Qy 241 LFTAVSHGHLDVCV 253
Db 504 IFLQ-----FIDCV 512
RESULT 8
MTRL_HUMAN
ID MTRL_HUMAN STANDARD; PRT; 662 AA.
AC Q136I3; Q9UBX6; Q9UEM0; Q9UOD5;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myotubularin-related protein 1 (Fragment).
GN MTMRI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99047532; PubMed=9828128;
RA Kioschis P., Wiemann S., Heiss N.S., Francis F., Goetz C., Poustka A.,
Taudien S., Platzter M., Wiehe T., Beckmann G., Weber J., Nordsiek G.,
Rosenthal A.;
RT "Genomic organization of a 225-kb region in Xq28 containing the gene
for X-linked myotubular myopathy (MTM1) and a related gene (MTMRI).";
RL Genomics 54:256-266(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA Kioschis P., Wiemann S., Francis F., Goetz C., Poustka A., Taudien S.,
Platzter M., Wiehe T., Beckmann G., Weber J., Nordsiek G.,
Rosenthal A.;
RT "Ancient genomic duplication within the myotubular myopathy locus
(MTM1) in human Xq28";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 111-607 FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buji-Bello A., Tentler D., Kretz C., Dahl N.,
Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
gene family from yeast to human";
RL Hum. Mol. Genet. 7:1703-1712(1998).
RN [4]
RP SEQUENCE OF 401-532 FROM N.A.
RX MEDLINE=96225444; PubMed=8640223;
RA Laporte J., Hu L.J., Kretz C., Mandel J.-L., Kioschis P., Coy J.,
Klauck S.M., Poustka A., Dahl N.;
RT "A gene mutated in X-linked myotubular myopathy defines a new
putative tyrosine phosphatase family conserved in yeast.";
RL Nat. Genet. 13:175-182(1996).
CC -1- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q136I3-1; Sequence=Displayed;
CC Name=1A;
CC IsoId=Q136I3-2; Sequence=VSP_005169, VSP_005170;
CC -1- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.

RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayaishizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: Contains 6 ANK repeats.
 CC -1- SIMILARITY: Contains 1 SOCS box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF398970; AAK97492.1; -;
 DR EMBL; AK008978; BAB24293.1; -;
 DR MGD; MGI:1916549; ASB9.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001496; SOCS.
 DR Pfam; PF00023; ank; 6.
 DR SMART; SM00248; ANK; 6.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50225; SOCS; 1.
 KW ANK repeat; Repeat.
 FT REPEAT 31 60 ANK 1.
 FT REPEAT 64 93 ANK 2.
 FT REPEAT 97 126 ANK 3.
 FT REPEAT 129 158 ANK 4.
 FT REPEAT 162 191 ANK 5.
 FT REPEAT 194 223 ANK 6.
 FT DOMAIN 236 290 SOCS BOX.
 FT CONFLICT 275 275 C -> S (IN REF. 2).
 SQ SEQUENCE 290 AA; 31672 MW; 25F99A086951B856 CRC64;

Query Match 10.7%; Score 250; DB 1; Length 290;
 Best Local Similarity 29.7%; Pred. No. 2e-12;
 Matches 88; Conservative 36; Mismatches 112; Indels 60; Gaps 10;

QY 170 LNAQVYNDSDYLDQLRQERYKRFINSRGV-----PGTPLRLAASYGHLSCLQVL 223
 DB 36 LKDAIHGCLLTNRNLSQ-----GMPVNIITADHVSPLHEACRLGHLSCASVL 84
 QY 224 LAHGADVDSLDVKAQTPFTAVSHGHLDCVRVLLLEAGSPGSIYNNCSPLVLTAAARDGAV 283
 DB 85 LSHGAQVNGMTDWRTPFNACVSGQDCVNLQHGATPHPET-ELASPIHEAAKRGV 143
 QY 284 AILQELLDHGAENAVKAKLPVWASNTAGSGPLYLAAYVGHLDPCFRLLLHGADEPDYNT 343
 DB 144 KCIESLAAGANTIDY-----NISHLGTPLYVACKNQVACAKKLLSGVS-----V 189
 QY 344 DQG-----LLARVPRPTLLEICLHNCPEPEYIQLLDGFIYPLSL-----LDIT 391
 DB 190 NQGGKGLDSPHVAVM-----SSVELVHLLMDFGANAQAKNADGKRPVDIV 235

QY 392 SODDKGIALLLQARATPRSLLSQVRLVVRALCQAGQPOAINQLDIPPLMISYIKH 447
 DB 236 PLESPLIQIFLQNEG-POSIRQLCRLIRKCF-GIRQHHKICELLPLDKRFLH 289
 RESULT 14
 ID AS11 HUMAN STANDARD; PRT; 323 AA.
 AC Q8WXH4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 11 (ASB-11).
 GN ASB11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kile B.T., Hilton D.J., Nicola N.A.;
 RT "SOCS box proteins.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Contains 6 ANK repeats.
 CC -1- SIMILARITY: Contains 1 SOCS box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF425642; AAL60519.1; -;
 DR Genew; HGNC:17186; ASB11.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001496; SOCS.
 DR Pfam; PF00023; ank; 6.
 DR SMART; SM00248; ANK; 6.
 DR PROSITE; PS50088; ANK_REPEAT; 6.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50225; SOCS; 1.
 KW ANK repeat; Repeat.
 FT REPEAT 64 93 ANK 1.
 FT REPEAT 97 126 ANK 2.
 FT REPEAT 130 159 ANK 3.
 FT REPEAT 162 191 ANK 4.
 FT REPEAT 195 224 ANK 5.
 FT REPEAT 227 256 ANK 6.
 FT DOMAIN 273 323 SOCS BOX.
 SQ SEQUENCE 323 AA; 35367 MW; F95A40184C5BBF4E CRC64;

Query Match 10.5%; Score 246; DB 1; Length 323;
 Best Local Similarity 29.9%; Pred. No. 4.7e-12;
 Matches 82; Conservative 44; Mismatches 100; Indels 48; Gaps 9;

QY 206 TPRLRLAASYGHLSCLQVLLAHGADVDSLDVKAQTPFTAVSHGHLDCVRVLLLEAGSPG 265
 DB 67 SPLHEAAQGRLLALKTLIAQGVNVLVTINRVSSLHEACRLGHVACAKLLENGAHVNG 126
 QY 266 SIYNNCSPLVLTAAARDGAVAILQELLDHGAENAVKAKLPWASNTAGSGPLYLAAYVGH 325
 DB 127 VTVHGATPLFNACSGSGAACVNVLLFEGAKAQLEVLH-----ASPIHEAVKRGHR 176
 QY 326 DCFRLLLLHGADEPDYN-----CTDQGLLARVPRPTLLEI-----CL 362
 DB 177 ECWEILLANNVNIHVEVQLGTFTPLVACTYQ-----RVCVKKLLLEAGSDVHGWLDTL 232
 QY 363 H---HNCPEPEYIQLLDGFIYPLSL-----SLDITS-ODDKGIALLLQARATPRSLLSQ 414
 DB 233 HAAARQSNVEVTHLLTDYGANLKRENAQKGSALDLAAPKSSVEQALL--REGPPALSOL 290

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:44:47 ; Search time 92 Seconds
(without alignments)
2154.142 Million cell updates/sec

Title: US-09-941-831a-20
Perfect score: 2343
Sequence: 1 MRESGWKLDPISDFGRMG1.....QAINQLDIPPLISYLKHQL 449

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DRV=xlsp
-Q=/cgn2_1/USPTO.spool_p/US09941831/runat_01122003_073528_7855/app_query.fasta_1.647
-DB=issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09941831@cgn_1_1_85/runat_01122003_073528_7855 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCUTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480	20.5	2019	4	US-09-302-769-24
2	264	11.3	3431	4	US-09-155-078-1
3	261.5	11.2	1206	4	US-09-620-312D-34
4	244.5	10.4	401	3	US-09-221-298-28
5	233.5	10.0	419	4	US-09-302-769-26
6	219	9.3	3498	4	US-09-350-982C-4
7	219	9.3	4512	4	US-09-350-982C-3
8	213.5	9.1	4431	3	US-09-196-387-7
9	213.5	9.1	4491	4	US-09-841-835-7
10	209	8.9	4657	3	US-09-196-387-9
11	209	8.9	4657	4	US-09-841-835-9
12	208.5	8.9	4134	3	US-09-196-387-1

13	208.5	8.9	4134	4	US-09-841-835-1	Sequence 1, Appli
14	207	8.8	1712	4	US-09-620-312D-1101	Sequence 1101, Ap
15	205	8.7	14770	4	US-09-220-132-30	Sequence 30, Appl
16	202	8.6	3454	3	US-09-082-089-1	Sequence 1, Appli
17	201.5	8.6	909	1	US-09-031-485-25	Sequence 25, Appl
18	201.5	8.6	909	1	US-09-031-485-26	Sequence 26, Appl
19	201.5	8.6	909	1	US-08-847-429A-25	Sequence 25, Appl
20	201.5	8.6	909	1	US-08-847-429A-26	Sequence 26, Appl
21	201.5	8.6	909	3	US-09-065-474-25	Sequence 25, Appl
22	201.5	8.6	909	3	US-09-065-474-26	Sequence 26, Appl
23	201.5	8.6	909	4	US-09-557-034-25	Sequence 25, Appl
24	201.5	8.6	909	4	US-09-557-034-26	Sequence 26, Appl
25	201.5	8.6	911	1	US-09-031-485-22	Sequence 22, Appl
26	201.5	8.6	911	1	US-09-031-485-24	Sequence 24, Appl
27	201.5	8.6	911	1	US-08-847-429A-22	Sequence 22, Appl
28	201.5	8.6	911	1	US-08-847-429A-24	Sequence 24, Appl
29	201.5	8.6	911	3	US-09-065-474-22	Sequence 22, Appl
30	201.5	8.6	911	3	US-09-065-474-24	Sequence 24, Appl
31	201.5	8.6	911	4	US-09-557-034-22	Sequence 22, Appl
32	201.5	8.6	911	4	US-09-557-034-24	Sequence 24, Appl
33	201.5	8.6	1056	3	US-09-065-474-138	Sequence 138, App
34	201.5	8.6	1056	3	US-09-065-474-140	Sequence 140, App
35	201.5	8.6	1056	4	US-09-557-034-138	Sequence 138, App
36	201.5	8.6	1056	4	US-09-557-034-140	Sequence 140, App
37	201.5	8.6	5235	1	US-09-031-485-35	Sequence 35, Appl
38	201.5	8.6	5235	1	US-09-031-485-36	Sequence 36, Appl
39	201.5	8.6	5235	1	US-08-847-429A-35	Sequence 35, Appl
40	201.5	8.6	5235	1	US-08-847-429A-36	Sequence 36, Appl
41	201.5	8.6	5235	3	US-09-065-474-35	Sequence 35, Appl
42	201.5	8.6	5235	3	US-09-065-474-36	Sequence 36, Appl
43	201.5	8.6	5235	4	US-09-557-034-35	Sequence 35, Appl
44	201.5	8.6	5235	4	US-09-557-034-36	Sequence 36, Appl
45	201.5	8.6	5503	1	US-09-031-485-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-302-769-24
; Sequence 24, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicolas A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 24
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (2000)
; OTHER INFORMATION: n is unsure
US-09-302-769-24

Alignment Scores:
Pred. No.: 3,98e-48 Length: 2019
Score: 480.00 Matches: 114
Percent Similarity: 57.7% Conservative: 58

Best Local Similarity: 38.26% Mismatches: 104
Query Match: 20.49% Indels: 22
DB: 4 Gaps: 8

US-09-941-831A-20 (1-449) x US-09-302-769-24 (1-2019)

```
QY 170 LeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGlu 189
Db 167 CTCATGATGCAGCTATGTAGGGACCTCCAGACCTCAGAACCTACTGCAAGAGGAG 226
QY 190 ArgTyrLysArgPheLeuAsnSerArgSerGlyTrp-----GlyValProGlyThr 206
Db 227 AGCTACCGAGCCGCGATCAATGAGAAGTCTGTCTGGTGTGCGGCTTCCCTGCACA 286
QY 207 ProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHis 226
Db 287 CCACTGAGGATCGAGCCACTGAGGCCATGGGAACCTGTGTGAGCTTCCTCATACGCAAA 346
QY 227 GlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSer 246
Db 347 GGGGCCGAGGTGACCTGTGGTGTGATGTCAGGGCAGACTCCCTGTATGGCTGTAGTG 406
QY 247 HisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySer 266
Db 407 AACGGGCACCTTGGAGAGCAGCTGAGATCCTTTTGGAGCTGTGTCTGATCCCAACGGCAGC 466
QY 267 IleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLeu 286
Db 467 CGGCACACCGCAGCACTCCTGTGTACCATGCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 526
QY 287 GlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu----- 302
Db 527 AAGGCTCTTATAGGTATGGGCGAGATGTTGATGTCAACCATCATCTGAAATCTGACACC 586
QY 303 ---ProValTrpAlaSerAniLeuAlaSer-----CysSerGlyProLeuTyrLeu 318
Db 587 CGGCCCCCTTTTTCACGGCGGCTAACCTCTCTTGGTGTCTGT-----CCTCTATACATC 640
QY 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuLeuLeuHisGlyAlaAspPro 338
Db 641 AGTGCTGCTTACCATAACCTTCACTGCTTCAAGTGTCTTGTGTGTGTGTGTGTGTGTGTGT 700
QY 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArg-----ProArgThr 356
Db 701 GACTTCATTCGAATGCCCTGTCAACCCAGAGGTTCTACAGGGATCCCTCGGGTGT 760
QY 357 LeuLeuGluLeuCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAsp 376
Db 761 GTCATGATGCTCTCTCGCCCATGGCTGTGAAGCAGCCTTCTGTAGTCTGTGTGTGTGTGTGT 820
QY 377 PheGlyAlaAsnIleTyrLeu-----ProSerLeuSerLeuAspLeuThrSerGln 393
Db 821 TTGGAGCCCAACCTGAACCTGTGTGAAGTGGGAATCCCTGGGCCCAGAGGAGGAGGAGGAG 880
QY 394 -----AspAspLysGlyIleAlaLeuLeuLeuGlnAlaArgAlaThrProArgSer 410
Db 881 AGAAGATGATCTCGAGGCTTGCAGGCTTTTAAAGAGCCAGAGATATTCACGAGCC 940
QY 411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 941 TTCTGATTTTGGCCGGGTGGCTGTGAGAAGAGCTTTGGCAATAC---CGACTGCAT 997
QY 431 AlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyrLeuLysHisGln 448
Db 998 CTGGTTCCCTCGCTGCGCTGCCAGACCCCAATAAAGAGTTTGTCTTATGAG 1051
```

RESULT 2

US-09-155-078-1
; Sequence 1, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra

; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Nikias
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078
; CURRENT FILING DATE: 1999-03-02
; EARLIER APPLICATION NUMBER: PCT/DE97/00592
; EARLIER FILING DATE: 1997-03-21
; EARLIER APPLICATION NUMBER: DE 196 11 234.6
; EARLIER FILING DATE: 1996-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatenIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3431
; TYPE: DNA
; ORGANISM: German-type microorganism & cell cul.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1863)
; US-09-155-078-1

Alignment Scores:
Pred. No.: 2,4e-21 Length: 3431
Score: 264.00 Matches: 80
Percent Similarity: 42.00% Conservative: 46
Best Local Similarity: 26.67% Mismatches: 106
Query Match: 11.27% Indels: 68
DB: 4 Gaps: 12

US-09-941-831A-20 (1-449) x US-09-155-078-1 (1-3431)

```
QY 5 GlyTrpLysLeuLeuAspProIleSerPheGlyArgMetGlyIleProAsnArgAsn 24
Db 541 GGATGCAGCTTTTACATCCAGTGGGAAGAAATACAGGAGCGAGGCTTGCCCAATCACCAT 600
QY 25 TrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyrProGluIle 44
Db 601 TCGAGAATAAATCTTTTATTAATAAGTGTCTGTAGCTCTGTGACACTTACCCCTCTCTTTG 660
QY 45 ValValProLysSerValThrLeuGlyThrValValGlySerSerLysPheArgSerLys 64
Db 661 GTGGTTCCGTATCGTCTGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 65 GluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAlaAlaIleCysArgCys 84
Db 721 AATCGAATTCAGTCTGTCTGATGGATTTCATCCAGAAAATAAGACGGTCTATTGTGGCTGC 780
QY 85 SerGlnProLeuSerGlyPheTyr---ThrArgCysValAspAspGluLeuLeuGlu 103
Db 781 AGTCAGCTCTTGTCCGGTATGATGCGGAACCGAAATAAAGATGATGAGAAATATCTCGAT 840
QY 104 AlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThrArgProLys 123
Db 841 GTTATCAGGAGACTTAATAACAAATTTCTAACTCACCATTATGATGCAAGACCCAGC 900
QY 124 Ile-----TrpHis 126
Db 901 GTAAATGCAGTGGCCCAACAGGCAAGGAGGATATGAAAGTGATGATGATGATGATGATGAT 960
QY 127 PheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMetAsnLeuMet----- 144
Db 961 AACGCCGAACCTTTCTTC-----TTAGACATTCATTAATATTCATGTTATGCGGAA 1011
QY 145 AspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu----- 159
Db 1012 TCTTTAAAAAAGTGAAGGACATGTTTATCTCTATGTAGAGATCTCATTTGTTGTCT 1071
QY 160 -----GluAspThrAspThrGluGlu-----LysGlnAlaLeuAsnGlnAlaValTyr 175
Db 1072 AGTTTGGAGTCTACTCTATTGGTTAGAACATATCAAGCTCGTTTTCAGAGGACCCATTCAA 1131
```

QY 176 AspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrIysArgPheille 195
Db 1132 GTAGCAGACAAA-----GTTTCTTCAGGGAAGAGTTTCAGTGTGTG 1173
QY 196 AnSerArgSerGlyTyrGlyValProGlyThrProLeuArgLeuAla----- 211
Db 1174 CATTCAGTGACGGATGGGACAGGACTGCTAGCTGACATCCTTGGCCATGCTGATGTG 1233
QY 212 -----AlaSer 213
Db 1234 GATAGCTTCTATAGGAGCAATGAAGGTTGCGAAATCTGCTACAAAAGAAATGGATAGT 1293
QY 214 TyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeu 233
Db 1294 TTTGGACATAAATTTGCAATCTCGAATA-----GGTCATGGT---GATAAAAACACACACC 1344
QY 234 AspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysVal 253
Db 1345 GATGCTGAC---CGTTCTCTCTATTTTCTCCAG-----TTTATTGATTGTGTG 1389

RESULT 3

US-09-620-312D-34
; Sequence 34, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghaast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 34
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)..(965)
US-09-620-312D-34

Alignment Scores:
Pred. No.: 8,05e-22 Length: 1206
Score: 261.50 Matches: 78
Percent Similarity: 44.64% Conservative: 47
Best Local Similarity: 27.86% Mismatches: 96
Query Match: 11.16% Indels: 59
DB: 4 Gaps: 7

US-09-941-831A-20 (1-449) x US-09-620-312D-34 (1-1206)

QY 206 ThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAla 225

Db 189 TCACCACCTACATGAAGCAGCAAGTCAAGTCCGCTTCTGTCTGAGAACATATTATCA 248
QY 226 HisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaVal 245
Db 249 CAGGTTTAAATGTAATGCAAGTAACTTAGACCATGTCACCCCATTCACGAAACCTGTC 308
QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
Db 309 CTGGAGATCAGCTGGCATGTCCCAAGTCTGTCTGGAGCAGGAGCTTAATGTAATGCA 368
QY 266 SerLeuTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLe 285
Db 369 ATCAGCATAGATGGCTGACTCCGTTATTCAACGATGCTCCCAAGGAGCTCAAGCTGT 428
QY 286 LeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTyr 305
Db 429 GCAGAGCTGCTTCTGGATGTGGTCCAAAGCCAGCTGGAGTCAATGCTTCCCA----- 482
QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeu 325
Db 483 -----TCCCAAGCATGAGCGCCGAGTAAGGTCAACAT 518
QY 326 AspCysPheArgLeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGln 345
Db 519 GAATGCTTGACATCTCATATCTCTGGGCGCATAGATGTTGAC----- 560
QY 346 GlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCys----- 361
Db 561 -----CAAGAAATTCCTCATTTGGGAACTCCTCTATGTAGTGTATGTCACAG 611
QY 361 ----- 361
Db 612 CAATTCATTCATCTGGAAGCTTTTATGCTGTGTGCTGACGTACAGAAAGGCAATAT 671
QY 362 -----LeuHis-----HisAsnCysGluProGluTyrIleGlnLeuLeu 374
Db 672 TGGGATCTCCCATTAACATGCTGCTCAACAATCCAGCACAGAAATTTGTAACATTACTG 731
QY 375 IleAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeu-----AspLeu 390
Db 732 CTAGAAATTTGGAGCAGATATCAATGCCAAAATACAGAGCTTCTCGACCTATAGATGTA 791
QY 391 ThrSerGlnAspAspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSer 410
Db 792 GCTAGCTTAGCAGTATGCTGGAAGAATATGCTTCAACATGAAGCTACCCCAAGCTCT 851
QY 411 LeuLeuSerGlnValArgLeuValValArgArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 852 CTTTACCAACTTTGCGGACTCTGTATCCGAAGCTACATA-----GGAACCAACA 902
QY 431 -----AlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyrLeuLysHisGln 448
Db 903 TTGCACCTTATCCCAACTCCAGTCCAGTCCAGCTTACTGAAGAATTTCTTACATGTCGA 962

RESULT 4

US-09-221-298-28
; Sequence 28, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Human
; FEATURE:

```
; NAME/KEY: modified_base
; LOCATION: (212)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-28

Alignment Scores:
Pred. No.: 1,45e-20 Length: 401
Score: 244.50 Matches: 46
Percent Similarity: 62.26% Conservative: 20
Best Local Similarity: 43.40% Mismatches: 39
Query Match: 10.44% Indels: 1
DB: 3 Gaps: 1

US-09-941-831A-20 (1-449) x US-09-221-298-28 (1-401)
QY 5 GlyTrpLysLeuLeuAspProIleSerAspPheGlyArgMetGlyLeuProAsnArgAsn 24
DB 71 GGATGGACAGTTTACAACTCCAGTGGGAAGATACAGGAGGCGAGGCTTGCCCAATCACCAT 130
QY 25 TrpThrIleThrAspAlaAsnArgAsnTyrGluileCysSerThrTyrPrProGluile 44
DB 131 TGGAGAAATAACTTTTATTATAAGTGTATAGCTCTGGACACTTACCCCTGCTCTTTTG 190
QY 45 ValValProLysSerValThrLeuGlyThrValValGlySerSerLysPheArgSerLys 64
DB 191 GTGGTCCGTATCGTCCTCANATGATGACCTCGGAGAGTTGCCAATTTTAGTCCCGA 250
QY 65 GluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAlaAlaIleCysArgCys 84
DB 251 AATCGAATTCAGTGTGTATGATTCATCCAGAAATAAGACGGTTCATTGTGCGTTGC 310
QY 85 SerGlnProLeuSerGlyPheTyr---ThrArgCysValAspGluLeuLeuGlu 103
DB 311 AGTCAGCCCTTGTCTCGGTATGAGTGGGAAACGAATAAAGATGATGAGAAATATCTCGAT 370
QY 104 AlaIleSerGlnThrAsn 109
DB 371 GTTATCAGGAGACTAAT 388

RESULT 5
US-09-302-769-26
; Sequence 26, Application US/09302769
; Patent No. 632317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 10976Z
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Human
US-09-302-769-26

Alignment Scores:
Pred. No.: 3,45e-19 Length: 419
Score: 233.50 Matches: 49
Percent Similarity: 65.35% Conservative: 17
Best Local Similarity: 48.51% Mismatches: 32
Query Match: 9.97% Indels: 3
```

```
DB: 4 Gaps: 1
US-09-941-831A-20 (1-449) x US-09-302-769-26 (1-419)
QY 166 GluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
DB 114 GAGGACACGAGGCTCCATGATGCAGCTTACGTGCGGACCTCCAGACCTCAGAGCCTA 173
QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyr-----Gly 202
DB 174 TTGCAAGAGGAGAGTACCGGAGCCGCATCAACGAGAAGTCTGTCTGGTGTGTGCTGG 233
QY 203 ValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeuGlnVal 222
DB 234 CTCCTCCACACCTTGGAAATCGGCCCATCGAGGCCATCGGAGCTGTGTGACCTTC 293
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
DB 294 CTCATCCGGAAGGGGCGGAGGTGGATCTGTGTGGACGTAAGAGACAGACGCCCTGTAT 353
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlnAlaGlyAlaSer 262
DB 354 GTGGTGTGTGAACGGGCACCTAGAGAGTACCCAGATCTTCTCGAAGCTGGCGGAC 413
QY 263 Pro 263
DB 414 CCC 416

RESULT 6
US-09-350-982C-4
; Sequence 4, Application US/09350982C
; Patent No. 6455290
; GENERAL INFORMATION:
; APPLICANT: Bertheisen, Jens
; APPLICANT: Toma, Salvatore
; APPLICANT: Isacchi, Antonella
; TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Relat
; TITLE OF INVENTION: Same
; FILE REFERENCE: PHRM-0043
; CURRENT APPLICATION NUMBER: US/09/350,982C
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 3498
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: PCR Primers
; NAME/KEY: CDS
; LOCATION: (1)..(3498)
; NAME/KEY: misc feature
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: (1102)..(1102)
; OTHER INFORMATION: n is any nucleic acid
; NAME/KEY: misc feature
; LOCATION: (2650)..(2650)
; OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-4

Alignment Scores:
Pred. No.: 7,8e-16 Length: 3498
Score: 219.00 Matches: 74
Percent Similarity: 40.98% Conservative: 26
Best Local Similarity: 30.33% Mismatches: 78
Query Match: 9.35% Indels: 66
DB: 4 Gaps: 7

US-09-941-831A-20 (1-449) x US-09-350-982C-4 (1-3498)
QY 189 GluArgTyrLysArgPhe-----IleAsnSerArgSerGlyTrpGlyVal 203
```

```
Db 112 GAACGAGTCAAGAGGCTGTGACGCTGAGAAGGTGAACAGCGCGGACACGCGGCGGCGAG 171
Qy 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
Db 172 AAATCCACCCGCTGCACTCGCGGAGGTTTGGCGGGAAGACGATGTTGAATATTG 231
Qy 224 LeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThr 243
Db 232 CTTCAAGATGGTCAAAATGCCAAGCAGCTGATGGGGCGCTTATTCTCTTCATAAT 291
Qy 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
Db 292 GCATGCTCTTTGGTCACTGAGTAGTCAATCTCTTTTGGCAGCATGGTGCAGACCC 351
Qy 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
Db 352 AATGCTCGAGATAATTGGAATTATATCTCTCTCATGAAGCTGCAATTAAAGAAAGATT 411
Qy 284 AlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuPro 303
Db 412 GATGTTTGATCTGTGCTGTACAGCATGGAGCTGAGCCCAACCATCCGAATACAGATGA 471
Qy 304 ValTrpAlaSerAsnIleAla----- 310
Db 472 AGGACAGCATTTGATTTAGCAGATCCATCTGCCAAGCAGTGCTTACTGCTGTAATAAAG 531
Qy 310 ----- 310
Db 532 AAAGATGAACCTTTAGAAAGTCCAGGAGTGGCAATGAAGAAATAATGATGGCTCTACTC 591
Qy 311 -----SerCys----- 310
Db 592 ACACCATTAATGTCACTGCCAGCGAGTGGGAGGAGTCACTCACTTACATTG 651
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 652 GCAGCAGGATATAACAGAGTAAGATTGTACAGCTGTACTGCAACATGGAGCT----- 705
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeu 358
Db 706 GATGTCATGCTAAAGATAAAGGTGATCTG-----GTACCA----- 741
Qy 359 GluIleCysLeuHisHisAsnCys-----GluProGlyTyrIleGlnLeuLeuIle 375
Db 742 -----TTACACAATGCCCTGTTCTTATGGTCATATGAAGTAAGTAAGTAAGTCTTGGTTC 792
Qy 376 AspPheGlyAla 379
Db 793 AAGCATGGTGCC 804
```

RESULT 7

US-09-350-982C-3

; Sequence 3, Application US/09350982C

; Patent No. 6455290

; GENERAL INFORMATION:

; APPLICANT: Berthelsen, Jens

; APPLICANT: Toma, Salvatore

; APPLICANT: Isacchi, Antonella

; TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Rela

; TITLE OF INVENTION: Same

; FILE REFERENCE: PHRM-0043

; CURRENT APPLICATION NUMBER: US/09/350,982C

; CURRENT FILING DATE: 1999-07-09

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 3

; TYPE: DNA

; LENGTH: 4512

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: PCR Primers

; NAME/KEY: misc feature

; LOCATION: (1124)...(1124)

```
; OTHER INFORMATION: n is any nucleic acid
; NAME/KEY: misc feature
; LOCATION: (2672)...(2672)
; OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-3
```

```
Alignment Scores:
Pred. No.: 1,21e-15 Length: 4512
Score: 219.00 Matches: 74
Percent Similarity: 40.98% Conservative: 26
Best Local Similarity: 30.33% Mismatches: 78
Query Match: 9.35% Indels: 66
DB: 4 Gaps: 7
```

US-09-941-831A-20 (1-449) x US-09-350-982C-3 (1-4512)

```
Qy 189 GluArgTyrLysArgPhe-----IleAsnSerArgSerGlyTrpGlyVal 203
Db 134 GAACGAGTCAAGAGGCTGTGACGCTGAGAAGGTGAACAGCGCGGACACGCGGCGGCGAG 193
Qy 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
Db 194 AAATCCACCCGCTGCACTCGCGGAGGTTTGGCGGGAAGACGATGTTGAATATTG 253
Qy 224 LeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThr 243
Db 254 CTTCAAGATGGTCAAAATGCCAAGCAGCTGATGGGGCGCTTATTCTCTTCATAAT 313
Qy 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
Db 314 GCATGCTCTTTTGGTCACTGCTGAGTAGTCAATCTCTTTTGGCAGCATGGTGCAGACCC 373
Qy 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
Db 374 AATGCTCGAGATAATTGGAATTATATCTCTCTCATGAAGCTGCAATTAAAGAAAGATT 433
Qy 284 AlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuPro 303
Db 434 GATGTTTGATCTGTGCTGTACAGCATGGAGCTGAGCCCAACCATCCGAATACAGATGA 493
Qy 304 ValTrpAlaSerAsnIleAla----- 310
Db 494 AGGACAGCATTTGATTTAGCAGATCCATCTGCCAAGCAGTGCTTACTGCTGTAATAAAG 553
Qy 310 ----- 310
Db 554 AAAGATGAACCTTTAGAAAGTCCAGGAGTGGCAATGAAGAAATAATGATGGCTCTACTC 613
Qy 311 -----SerCys----- 310
Db 614 ACACCATTAATGTCACTGCCAGCGAGTGGCAAGTGAAGTCACTCACTTACATTG 673
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 674 GCAGCAGGATATAACAGAGTAAGATTGTACAGCTGTACTGCAACATGGAGCT----- 727
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeu 358
Db 728 GATGTCATGCTAAAGATAAAGGTGATCTG-----GTACCA----- 763
Qy 359 GluIleCysLeuHisHisAsnCys-----GluProGlyTyrIleGlnLeuLeuIle 375
Db 764 -----TTACACAATGCCCTGTTCTTATGGTCATATGAAGTAAGTAAGTAAGTCTTGGTTC 814
Qy 376 AspPheGlyAla 379
Db 815 AAGCATGGTGCC 826
```

RESULT 8

US-09-196-387-7

; Sequence 7, Application US/09196387

; Patent No. 6277613

; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/196,387
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/095,225
 ; FILING DATE: June 10, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4491 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 6..2027
 ; US-09-196-387-7

Alignment Scores:
 Pred. No.: 5,63e-15 Length: 4491
 Score: 213.50 Matches: 102
 Percent Similarity: 37.50% Conservative: 42
 Best Local Similarity: 26.56% Mismatches: 143
 Query Match: 9.11% Indels: 97
 DB: 3 Gaps: 15

US-09-941-831A-20 (1-449) x US-09-196-387-7 (1-4491)

QY 136 LeuGlnLeuAlaLysMetAsnLeu-----MetAspIleThrLys 148
 Db 2205 TTGCAGCTCTCAAAATGTGAATGTAGAGACTTAGAGGGCGGCAATCCAGGCGCTTACA 2264
 QY 149 IlePheSerLeuLeuGlnProAspLysGluGluAspThrAspThrGluGluLysGln 168
 Db 2265 CTTTCGAGCAGCGGTACACAGAGAGTACACCTATACGTACTTCTGATGTTGATTATCGA--- 2321
 QY 169 AlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGln 188
 Db 2322 ---CTTTAGAGCATCTAAAGCTGAGACTTGGAAACTGTGAGCAACTTTGCAGCTCT 2378
 QY 189 GluArgTyrLysArgPheIleAsnSerArgSerGlyTyrPblyValProGlyThrProLeu 208
 Db 2379 CAAAAT-----GTGAATTGTAGAGACTTAGAGGGCGGCGATTCCAGGCCCTTA 2426
 QY 209 ArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAla 228

Db 2427 CACTTCGCAGCAGGCTACAAACCCGCTGTCTGTGTAGACTACCTGTCACACCCAGGTGCC 2486
 QY 229 AspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGly 248
 Db 2487 GATGTCCATGCCAAAGACAAGGGTGGCTGGTCCCTTCATATATCCCTGTTTCATATGGA 2546
 QY 249 HisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyr 268
 Db 2547 CACTATGAGTGGCTGAGCTTTTAGTAAGGCATGGGCTTCTGTCAATGTGGGGACTTA 2606
 QY 269 AsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGlu 288
 Db 2607 TGGAAATTTACCCCTCTCCATGAAGCAGCAGCTAAAGGAAAGTATGAAATCTGCAAGCTC 2666
 QY 289 LeuLeuAspHisGlyAlaGluAlaAsnValLys----- 299
 Db 2667 CTTTAAACATGGAGCAGATCCAACTAAAGAACAGAGATGGAATACACCTTTGGAT 2726
 QY 299 ----- 299
 Db 2727 TTGGTAAAGGAAGGAGACACAGATATTAGGACTTACTGAAAGGGGATGCTGCTTTGTTG 2786
 QY 300 -----AlaLysLeuProValTrpAla-----SerAsnIle 309
 Db 2787 GATGTCGCCAAGAAGGGCTGCCTGGCAAGAGTGCAGAACTCTGTACCCAGAGAAATATC 2846
 QY 310 AlaSerCys-----SerGlyProLeuTyrLeuAlaAlaValTyr 322
 Db 2847 ---AACTGCAGACACCCAGGGCAGAAATTCACCCCTCTGCACCTGCAGCAGGCTAT 2903
 QY 323 GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAsp----- 337
 Db 2904 AATAACCTGGAAGTAGCTGAATATCTTCTAGAGCATGGAGCTGATGTTAATGCCAGGAC 2963
 QY 338 -----ProAspTyrAsnCysThrAspGlnGly-----LeuLeuAla 349
 Db 2964 AAGGGTGGTTTAATTCCTCTTCAATATGCGGCATCTTATGGGCATGTTGACATAGCGGCT 3023
 QY 350 ArgValProArgProArgThrLeuLeuGluIleCys-----Leu 362
 Db 3024 TTATTGATAAAATACACACGCTGTGTAATGCAACAGATAAGTGGCGGTTTACTCCCTC 3083
 QY 363 HisHisAsnCysGluProGluTyrIleGln-----LeuLeuIleAspPheGlyAla 379
 Db 3084 CATGAAGCAGCCAGAAAGGAGGAGCGAGCTGTGCGCCCTCTCTCTAGCGCATGTTGCA 3143
 QY 380 AsnIleTyrLeuProSerLeuSer-----LeuAspLeuThrSer 392
 Db 3144 GAC-----CCACCATGAAGAACCCAGGAAGCCAGACGCTCTGGATCTGGAACA 3194
 QY 393 GlnAspAspLysGlyIleAlaLeuLeuGlnAla-----ArgAlaThrProArg 409
 Db 3195 GCTGACGATATCAGA---GCTTTGCTGATAGATGCATGCCGCCAGAGGCTTACTTACC 3251
 QY 410 SerLeuLeuSerGlnValArgLeuValValArgAlaLeuGlnAlaGlyGlnPro 429
 Db 3252 TGTTTAAACCTCAGGCTACTGTAGTGGTGGCTCTCTGATCTCACCAGCATCCACCCCC 3311
 QY 430 GlnAlaIleAsn 433
 Db 3312 TCCTGCGCTCTCG 3323
 RESULT 9
 US-09-841-835-7
 ; Sequence 7, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:
 ; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 4491 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 6..2027

US-09-841-835-7

Alignment Scores:

Pred. No.: 5.63e-15 Length: 4491
Score: 213.50 Matches: 102
Percent Similarity: 37.50% Conservative: 42
Best Local Similarity: 26.56% Mismatches: 143
Query Match: 9.11% Indels: 97
DB: 4 Gaps: 15

US-09-941-831A-20 (1-449) x US-09-841-835-7 (1-4491)

QY 136 LeuGlnLeuAlaLysMetAsnLeu-----MetAspIleThrLys 148
DB 2205 TTGCAGCTCTCAAAATGTGNAITGTAGAGCTTAGAGGGCGGCATTCACAGCCCTTACA 2264
QY 149 IlePheSerLeuLeuGlnProAspLysGluGluAspThrAspThrGluGluLysGln 168
DB 2265 CTTGCGAGCAGGCTACACAGAGTACACCTATACGTACTTCTGATGTTGATTATCA--- 2321
QY 169 AlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGln 188
DB 2322 ---CTCTTAGAGGCATCTAAAGCTGGAGACTTGGAAACTGTGAAGCAACTTTCAGCTCT 2378
QY 189 GluArgTyrLysArgPheIleAsnSerArgSerGlyTrpGlyValProGlyThrProLeu 208
DB 2379 CAAAT-----GTGAATGTAGAGACTTAGAGGGCGGCATTCACAGCCCTTA 2426
QY 209 ArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAla 228
DB 2427 CACTTCGACAGGCTACACCGCGTCTGTTGTAGAGTACCTGTACACAGCGTGCC 2486
QY 229 AspValAspSerLeuLeuValLysAlaGlnThrProLeuPheThrAlaValSerHisGly 248
DB 2487 GATGTCATCCCAAGACAGAGGGTGGCTTGGTCCCTTTCATTAATGCCTGTTTCATATGGA 2546

QY 249 HisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyr 268
DB 2547 CACTATGAGTGGCTGAGCTTTTACTAGGCGATGGGCTTCTGTCAATGTGGGACTTA 2606
QY 269 AsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGlu 288
DB 2607 TGGAAATTTACCCCTCTCCATGAAGCAGCAGCTAAAGGAAGTATGAAATTCGACGCTC 2666
QY 289 LeuLeuAspHisGlyAlaGluAlaAsnValLys----- 299
DB 2667 CTTTTAAACATGGAGCAGATCCAACTAAAGAACACAGAGATGGAAATACACCTTTGGAT 2726
QY 299 ----- 299
DB 2727 TTGGTAAAGGAGAGACACAGATATTCAGGACTTACTGAAAGGGGATCTGCTTTGTTG 2786
QY 300 -----AlaLysLeuProValTrpAla-----SerAsnIle 309
DB 2787 GATGCTGCCAAGAGGCTGCTGCGCAAGAGTGCAGAGCTCTGTACCCAGAGATATC 2846
QY 310 AlaSerCys-----SerGlyProLeuTyrLeuAlaAlaValTyr 322
DB 2847 ---AACTGCAGAGACACCCAGGCGCAAAATTCACCCCTCTGCACCTGCAGAGGCTAT 2903
QY 323 GlyHisLeuAspCysPheArgLeuLeuHisGlyAlaAsp----- 337
DB 2904 AATAACCTGGAAGTAGCTGAATATCTTCTAGAGCATGGAGCTGATGTTAATGCCAGGAC 2963
QY 338 -----ProAspTyrAsnCysThrAspGlnGly-----LeuLeuAla 349
DB 2964 AAGGGTGGTTAAATTCCTCTTCATTAATGCGCATCTTATGGGCATGTTGACATAGCGCT 3023
QY 350 ArgValProArgProArgThrLeuLeuGluIleCys-----Leu 362
DB 3024 TTATTGATAAAATACAAACACGTGTGTAATCAACAGATAGTGGCGGTTTACTCCCTC 3083
QY 363 HisHisAsnCysGluProGluTyrIleGln-----LeuLeuIleAspPheGlyAla 379
DB 3084 CATGAAGCAGCCAGAGAGGAGGAGCGAGCTGTGGCCCTCTCTCTAGCGCATGTTGCA 3143
QY 380 AsnIleTyrLeuProSerLeuSer-----LeuAspLeuThrSer 392
DB 3144 GAC-----CCACCATGAAGAACACAGGAAGGCGCAGCGCTCTGTGATCTGGCAACA 3194
QY 393 GlnAspAspLysGlyIleAlaLeuLeuGlnAla-----ArgAlaThrProArg 409
DB 3195 GCTGACGATATCAGA---GCTTTGCTGATAGATGCCATGCCCCAGAGCGCTTACCTACC 3251
QY 410 SerLeuLeuSerGlnValArgLeuValArgAlaLeuCysGlnAlaGlyGlnPro 429
DB 3252 TGTTTTAAACCTCAGGCTACTGTAGTGGCTCTCTGATCTCACCAGCATCCACCCCC 3311
QY 430 GlnAlaIleAsn 433
DB 3312 TCCTGCTCTCG 3323

RESULT 10

US-09-196-387-9
Sequence 9, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601


```
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4657 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 6..2855
US-09-841-835-9

Alignment Scores:
Pred. No.: 2,12e-14 Length: 4657
Score: 209.00 Matches: 93
Percent Similarity: 36.28% Conservative: 30
Best Local Similarity: 27.43% Mismatches: 122
Query Match: 8.92% Indels: 94
DB: 4 Gaps: 11

US-09-941-831A-20 (1-449) x US-09-841-835-9 (1-4657)
QY 139 AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
DB 1887 GCACAGATGGCAATGAACAGTGCAGAGATCTCAGTGACAGTACACTATACGT--- 1943
QY 159 GluGluAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
DB 1944 ACTTCGTGATTGAT-----TATCGACTCTTAGAGGCATCTAAAGCTGGAGAC 1991
QY 179 SerTyrThrLeuAspGlnLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
DB 1992 TTGGAACCTGTGAAGCACTTTGCAGCTCTCAAAAT-----GTGAATTGTAGA 2039
QY 199 SerGlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
DB 2040 GACTTAGAGGGCGGCAATCCAGCCCTTACACTTCGAGCAGGCTACACCGCGTGT 2099
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
DB 2100 GTTGTAGAGTACTGTCTACACACGCGTCCGATGTCATGCCAAAGACAAAGGTGGCTG 2159
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
DB 2160 GTGCCCTTTCAATATCCCTGTTTCATATGGACACTATGAGGTGGCTGAGCTTTTGTAG 2219
QY 259 AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
DB 2220 CATGGGGCTTCTGCAATGTGGCGGACTTATGGAAATTTACCCCTCTCCATGAGCAGCA 2279
QY 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
DB 2280 GCTAAAGGAAAGTATCAATCTGCAAGCTCTTTTAAACATGGAGCAGATCCAACTAA 2339
QY 299 Lys----- 299
DB 2340 AAGAACAGAGATGGAATAACCTTTGGATTGTTGTAAGGAGGAGACACAGATATTCAG 2399
QY 300 -----AlaLysLeuProValTyrPala--- 306
DB 2400 GACTTACTGAAGGGGATGCTGTTTGGATGCTGCCAAGAGGGTGCCTGGCAAGA 2459
QY 307 -----SerAsnIleAlaSerCys----- 312
DB 2460 GTGCAGAGCTCTGTACCCAGAGATATC---AAGTGCAGACACCCAGGCGCAGAAAT 2516
QY 313 SerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeu 332
DB 2517 TCAACCCCTCTGCACCTGGCAGCAGGCTATATAACCTGGAAGTAGCTGAATATCTCTTA 2576
```

333 LeuHisGlyAlaAsp-----ProAspTyrAsnCys 342
2577 GAGCATGGAGCTGATGTTAATGCCAGGACAGGCTGTTTAAATTCCTTCTTCAATATGCG 2636
343 ThrAspGln---GlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeuCys 361
2637 GCATCTTATGGGGCTGCTGCTGCAAGAGTGCAGAAG----- 2672
362 LeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381
2673 -----CTCTGTACCCAGAGATATCACTCCAGACACACCAGGCGCAGAAATTC 2723
382 -----TyrLeuProSer 385
2724 ACCCTCTGCACCTGGCAGCAGGCTAATAACCTGGAAGTAGCTGAATATCTTCTAGAG 2783
386 LeuSerLeuAspLeuThrSerGlnAspLysGlyIleAlaLeuLeuLeuGlnAla 404
2784 CATGGAGCTGATGTTAATGCCAGGACAGGCTGTTTAAATTCCTTCTTCAATATGCG 2840

RESULT 12
US-09-196-387-1
; Sequence 1, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-09-196-387-1

Alignment Scores:
Pred. No.: 1,99e-14 Length: 4134
Score: 208.50 Matches: 100
Percent Similarity: 37.70% Conservative: 41
Best Local Similarity: 26.74% Mismatches: 140
Query Match: 8.90% Indels: 93

DB: 3 Caps: 15

US-09-941-831A-20 (1-449) x US-09-196-387-1 (1-4134)

Qy 139 AlaLysMetAsnLeuMetAspPheSerLeuLeuGlnProAspLysGlu 158
 Db 1887 GCACAGATGGCAATGAGCAGATCTCTGAGTGAGAGTACACCTATACGT--- 1943

Qy 159 GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
 Db 1944 ACTTCTGATGTTGAT-----TATCGACCTTTAGAGGCATCTAAAGCTGGAGAC 1991

Qy 179 SerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
 Db 1992 TTGGAACCTGTGAAGCAACTTTCAGCTCTCAAAAT-----GTGAATTTGAGA 2039

Qy 199 SerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
 Db 2040 GACTTAGAGCGCGGCAATTCACGCCCTTACACTTCGCAGCAGGCTACAAACCGGTGTCT 2099

Qy 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValValAlaGln 238
 Db 2100 GTTGAGTAGACTGCTACACACGCGTGCAGATGTCATGCCAACAGACAGGGTGGCTTG 2159

Qy 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
 Db 2160 GTGCCCTTCATAATGCCTGTTTCATATGACACTATGAGTGGCTGAGCTTTTAGTAGG 2219

Qy 259 AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
 Db 2220 CATGGGCGCTCTCTCAATGTGGCGGACTTATGGAAATTTACCCCTCTCCATGAAGCAGCA 2279

Qy 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
 Db 2280 GCTAAAGGAAGTAGTAATCTCGAAGCTCTTTTAAACATGGAGCAGATCCCACTAAA 2339

Qy 299 Lys-----SerAsnIleAlaSerCys----- 299
 Db 2340 AAGACAGAGATGGAAATACACCTTTGGATTGGTAAAGGAGGAGACACAGATATTTCAG 2399

Qy 300 -----AlaLysLeuProValTrpAla--- 306
 Db 2400 GACTTACTGAAGGGGATGCTGCTTTGGTGGATGCTGCCAAGAGGGGCTGCTGGCAAGA 2459

Qy 307 -----SerAsnIleAlaSerCys----- 312
 Db 2460 GTGCAGAAGCTCTGTACCCAGAGATATC---ACTGCAGAGACACCCAGGGCAGAAAT 2516

Qy 313 SerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeu 332
 Db 2517 TCAACCCCTCTGCACCTGCAGCAGCTATAATAACCTGGAAGTAGCTGAATATCTTCTA 2576

Qy 333 LeuHisGlyAlaAsp-----ProAspTyrAsnCys 342
 Db 2577 GAGCATGGAGCTGATGTTAATGCCAGGACAGGGTGGTTTAAATTCCTCTTCATAATGCG 2636

Qy 343 ThrAspGlnGly-----LeuLeuAlaArgValProArgProArgThrLeuLeuGlu 359
 Db 2637 GCATCTATGGGCATGTTGCATAGGCGCTTATTGATAAAATACACACGTGTGTAAT 2696

Qy 360 IleCys-----LeuHisHisAsnCysGluProGluTyrIleGln 372
 Db 2697 GCAACAGATAAGTGGCGGTTTACTCCCTCCATGAAGCAGCCCAAGGAAGGAGCAGCAG 2756

Qy 373 -----LeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSerLeuSer----- 387
 Db 2757 CTGTGCGCCCTCTCTCGCATGGTGGCAGAC-----CCACCATGAAGAACCCAG 2807

Qy 388 -----LeuAspLeuThrSerGlnAspLysGlyIleAlaLeuLeuLeu 402
 Db 2808 GAAGGCCAGACCGCTCTGGATCTGGCAACAGCAGCTGACGATATCAGA---GCTTCTGCTGATA 2864

Qy 403 GlnAla-----ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValVal 419

```
Db 1992 TTGGAACCTGTGAAGCAACTTTGACGCTCTCAAAAT-:::|||||:::
Qy 199 SerGlyTrpGlyValProGlyThrProLeuArgLeuAlaSerTyrGlyHisLeuSer 218
Db 2040 GACTTAGAGGGCGGCGGATTCACCGCCCTTACACTTCGCGACAGGCTACACCGCGTGTCT 2099
Qy 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
Db 2100 GTTGTAGAGTACTGTACACACGGTGGCGATGTCATGCCAAGACAGGGTGGCTTG 2159
Qy 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
Db 2160 GTGCCCTTTCATAATGCTTTCATATGACACATGATGAGTGGCTGAGCTTTAGTAAGG 2219
Qy 259 AlaGlyAlaSerProGlySerGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
Db 2220 CATGGGGCTTCTGTCATGTGGCGGACTTATGGAAATTTACCCCTCTCCATGAAGCAGCA 2279
Qy 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
Db 2280 GCTAAAGGAAGTATGAATCTCAAGCTCTTTTAAACATGGACGACATCCAACTAA 2339
Qy 299 Lys----- 299
Db 2340 AAGAACAGAGATGGAATACACCTTTGGATTTGGTAAAGGAAGGACACAGATATTTCAG 2399
Qy 300 -----AlaLysLeuProValTrpAla----- 306
Db 2400 GACTTACTGAAGGGGATGCTGCTTTGTTGGATGCTGCCAAGAGGGCTGCTGGCAAGA 2459
Qy 307 -----SerAsnIleAlaSerCys----- 312
Db 2460 GTCCAGAGCTGTGTACCCAGAGATATC---AATGACAGACACCCAGGCGCAGAAAT 2516
Qy 313 SerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeu 332
Db 2517 TCAACCCCTCTGCACCTGGCAGCAGGCTATAATACTTGAAGTAGCTGAATATCTTTCTA 2576
Qy 333 LeuHisGlyAlaAsp-----ProAspTyrAsnCys 342
Db 2577 GAGCATGGAGCTGATGTTAATGCCAGGACAGGGTGGTTTAAATCTCTCTATATGGC 2636
Qy 343 ThrAspGlnGly-----LeuLeuAlaArgValProArgProArgThrLeuLeuGlu 359
Db 2637 GCATCTTATGGCATGTTGACATAGCGGCTTTATTGATAAAATACACACGCTGTGTAAT 2696
Qy 360 IleCys-----LeuHisAsnCysGluProGluTyrIleGln 372
Db 2697 GCAACAGATAAGTGGGGTTTACTCCCTTCATGAAGACCCAGAAAGGAGGACGACG 2756
Qy 373 -----LeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSerLeuSer----- 387
Db 2757 CTGTGGCCCTCTCTCTAGCGCATGGTGGCAGAC-----CCACCATGAAGAACCCAG 2807
Qy 388 -----LeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuLeu 402
Db 2808 GAAGGCCAGACGCTCTGGATCTGGCAACAGCTGACGATATCAGA---GCTTTGCTGGATA 2864
Qy 403 GlnAla-----ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValVal 419
Db 2865 GATGCCATGCCCCAGAGGCTTACCTACTCTGTTTAAACCTCAGGCTACTGTAGTAGT 2924
Qy 420 ArgArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsn 433
Db 2925 GCCTCTCTGATCTCACAGCATCCACCCCTCTCTGCTCTCG 2966
```

RESULT 14

```
US-09-620-312D-1101
; Sequence 1101, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
```

```
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1101
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (551)..(1672)
US-09-620-312D-1101
```

Alignment Scores:

```
Pred. No.: 6,68e-15 Length: 1712
Score: 207.00 Matches: 92
Percent Similarity: 42.90% Conservative: 44
Best Local Similarity: 25.02% Mismatches: 111
Query Match: 8.83% Indels: 70
DB: 4 Gaps: 14
```

US-09-941-831A-20 (1-449) x US-09-620-312D-1101 (1-1712)

```
Qy 157 LysGluGluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAsp 176
Db 208 CGCCAGAGTCAAGACTGGGGCAAGAGTGACGAGAGGCTGCTA---CAAGCCGTGGAAGAAC 264
Qy 177 AsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsn 196
Db 265 AACGATGACCTCGGGTGGCCCGCTCATCGCCCGCAAG----- 303
Qy 197 SerArgSerGlyTrpGlyValProGlyThrProLeu----- 208
Db 304 -----GGCTGTGTGCCACGACAGCTAGACCCCGAGGGCAAGTCCGCGTTC 348
Qy 209 ArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAla 228
Db 349 CACCTGGGGCCATCGGGGTGGCGGCGGCTGCTGTGGAGGTGATGATGATCATGGCAGC 408
Qy 229 AspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGly 248
Db 409 AATGTGATGAGCGGACGCGGCGGAGTTACAAATGCCCTCCACCTGGCCGCCAAATACGGG 468
Qy 249 HisLeuAspCysValArgValLeuLeuGluAlaGly-Ala-----Se 262
Db 469 CACCCACAGTGTGTGAAGCAACTACTGACGGCTTCTGCTGGTGGACCTCGTGGACAGC 528
Qy 262 rProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAl 282
Db 529 AGCGGGTGGACTGCCCTTACACCATGACGGGTGGATTGCCCTAGCAGCGGCTGGTGGCTG 598
```

```
QY 282 aValAlaIleLeuGlnGluLeuAspHisGlyAlaGluAlaAsnVallysAlaIle 302
Db 589 TCCTCTCTGCTCAGAGGTGCTCTCTCTTTAAGGCACATCTAAACCCCAAGATCGG-- 646
QY 302 uProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaValTy 322
Db 647 -----TCAGGCGCAACACACCCCTCATTTATAGCAGCTCAGAT 681
QY 322 rGlyHisLeuAspCysPheArgLeuLeuHisGlyAlaAspProAspTyrAsnCy 342
Db 682 GTGTCAACAGACTGTGCGCTCTCTCTACTGCAAGAGGGCT-----GCCGC 729
QY 342 sThrAspGlnGlyLeuAlaArgValProArgThrLeuLeuGluIleCysLe 362
Db 730 GAACGATCAGACCTCAAGCCAGGAGCG-----GCCCTGATGTGCGCTGT-- 775
QY 362 uHisAsnCyGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTy 382
Db 776 ----GAGGGGCGCCAGCCCGCAACAGTGGAGGTCTCTGCTCAGGCGGAGCCAG---- 826
QY 382 rLeuProSerLeuSerLeuAspLeuThrSerGlnAsp----- 394
Db 827 ----CCGGGATCACC-----GATGCGTGGGGCAGGAGCGGCTCACTATGGCGCCCTGGC 879
QY 395 ----AspLysGlyIleAlaLeuLeuGln-----AlaArg-AlaThrProArgSerL 411
Db 880 GGGGGACAACTACTCTGCTCAGCTTCTGCAAGAGCGGCGCCAGCCCTCCGCCAG 939
QY 411 euLeuSerGlnValArgLeuValValArg-----ArgAlaLeuCysGlnAla---- 426
Db 940 CGCCCTCAGACAGGATGATTTCAGGCGAGGGGTCTCTCAGAACTCTATGTCCAGCCATGG 999
QY 427 -----GlyGlnProGlnAlaIleAsnGlnLeuAspIleProPro 439
Db 1000 AAAGCAGGGGGGCCCCCAAGAGCGGAGGCGCTCCACCTCCCGCCA 1046

RESULT 15
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 14770
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-30
```

```
Alignment Scores:
Pred. No.: 4,75e-13 Length: 14770
Score: 205.00 Matches: 63
Percent Similarity: 45.42% Conservative: 46
Best Local Similarity: 26.25% Mismatches: 93
Query Match: 8.75% Indels: 38
DB: 4 Gaps: 9
```

US-09-941-831A-20 (1-449) x US-09-220-132-30 (1-14770)

```
QY 206 ThrProLeuArgLeuAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
Db 1198 TCTCCATGCACATGGCCCAACACAAAGGGGATCATTTAACTGGCTCCAGCTTCTCTCCAG 1257
```

```
QY 226 HisGlyAlaAspValAspSerLeuAspVallysAlaGlnThrProLeuPheThrAlaVal 245
Db 1258 CATATGTATCCCGTGGATGATGTACCATGACTACCTACCTACCTACCTACCTGGCTGCC 1317
QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
Db 1318 CACTGTGGCCATTACAAAGTTGCCAAGTTCCTCTGGATAAGAAAGCTAACCCCAATGCC 1377
QY 266 SerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaAlaIle 285
Db 1378 AAAGCCCTGAATGGCTTTACCCCTCTTCATATCGCCTGCAAGAGAATCGAATTAAGTA 1437
QY 286 LeuGlnGluLeuAspHisGlyAlaGluAlaAsnVallysAlaIleCysLeuProValTrp 305
Db 1438 ATGGAACTCTCTTGAACACCGT-----GCATCATCAAGCT-----GTAACC 1482
QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaValTyrGlyHisLeu 325
Db 1483 GAGTCGGGCTTACC-----CCAATCCATGTTGCTGCTTCATGGGCATGTA 1530
QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGln 345
Db 1531 AATATTTGTATCAACTAATGCATCATGAGCCTCACCA-----AACACCACCAAT--- 1581
QY 346 GlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHisAsn 365
Db 1582 -----GTGAGAGGAGAAACAGCAGCTGCACATGGCAGCTCGCTCCGCG 1623
QY 366 CysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSer 385
Db 1624 ---CAAGCTGAAGTTGTGCGGTATCTGTTACAAAGCAGGAGCTCAGGTA----- 1668
QY 386 LeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuGlnAlaArg 405
Db 1669 -----GAAGCTAAAGCTAAGGATGACCAA----- 1692
QY 406 AlaThrProArgSerLeuSerGlnValArgLeu-----ValValArgArgAlaLeu 423
Db 1693 ---ACACCACTCCACATTTTCAGCCCGAGCTGGGGAAGCAGACATAGTACACAGCTGTTG 1749
QY 424 CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSer 443
Db 1750 CAGCAAGGGGGCATCTCCAAATGCAGCCCAACTTCTGGGTACACCCCACTTCCCTTTCC 1809
```

Search completed: December 1, 2003, 09:51:26
Job time : 119 secs

RESULT 5

US-09-155-078-2

Sequence 2, Application US/09155078

Patent No. 6312688

GENERAL INFORMATION:

APPLICANT: POUSTKA, Annemarie

APPLICANT: KIOSCHIS, Petra

APPLICANT: LAPORTE, Jocelyn

APPLICANT: HU, Ling Jia

APPLICANT: MANDEL, Jean Louis

APPLICANT: DAHL, Niklas

TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN

FILE REFERENCE: 012627-010

CURRENT APPLICATION NUMBER: US/09/155,078

CURRENT FILING DATE: 1999-03-02

EARLIER APPLICATION NUMBER: PCT/DE97/00592

EARLIER FILING DATE: 1997-03-21

EARLIER APPLICATION NUMBER: DE 196 11 234.6

EARLIER FILING DATE: 1996-03-21

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 621

TYPE: PRT

ORGANISM: German-type microorganism & cell cul.

US-09-155-078-2

Query Match

Best Local Similarity 11.3%; Score 264; DB 4; Length 621;

Matches 80; Conservative 46; Mismatches 106; Indels 68; Gaps 12;

QY 5

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 65

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 124

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 301

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 160

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 358

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 212

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 412

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

QY 463

181 GWTVNPEVRYRQGLNHHWRTFKVCYELCDTYPALLVVPYRASDDDLRRVATFRSR 240

RESULT 6

US-09-350-982C-5

Sequence 5, Application US/09350982C

Patent No. 6455290

GENERAL INFORMATION:

APPLICANT: Berthelsen, Jens

APPLICANT: Toma, Salvatore

APPLICANT: Isacchi, Antonella

TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Rel

FILE REFERENCE: PHRM-0043

CURRENT APPLICATION NUMBER: US/09/350,982C

CURRENT FILING DATE: 1999-07-09

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 1166

TYPE: PRT

ORGANISM: Artificial

US-09-350-982C-5

Sequence 5, Application US/09350982C

Patent No. 6455290

GENERAL INFORMATION:

APPLICANT: Berthelsen, Jens

APPLICANT: Toma, Salvatore

APPLICANT: Isacchi, Antonella

TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Rel

FILE REFERENCE: PHRM-0043

CURRENT APPLICATION NUMBER: US/09/350,982C

CURRENT FILING DATE: 1999-07-09

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 1166

TYPE: PRT

ORGANISM: Artificial

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Xaa is any amino acid

NAME/KEY: misc feature

LOCATION: (1102)..(1102)

OTHER INFORMATION: n is any nucleic acid

NAME/KEY: misc feature

LOCATION: (2650)..(2650)

OTHER INFORMATION: n is any nucleic acid

US-09-350-982C-5

Query Match

Best Local Similarity 9.3%; Score 219; DB 4; Length 1166;

Matches 74; Conservative 26; Mismatches 78; Indels 66; Gaps 7;

QY 189

189 ERYKRF-----INSRSGMGVPTGTPRLRLAASGYHLSCLQVLLAHGADVDSLDVKAQTPLT 243

DB 38

38 ERVKELVTPPEKVNKRDTAGRKSTPLHLAAGFGKRVQVVEYLLQNGANVQARDGGGLPLHN 97

QY 244

244 AVSHGHLCVRLVLEAGASPGSIYNNCSPLVLTARDGAVAILQELLDHGAENVKAKLP 303

DB 98

98 ACSFGHAEVNVLLRHGADPNARDNNWYTPHLHAAIKGKIDVICVILLOHGAEPTRTGTG 157

QY 304

304 VNASNIA-----SC-----SGPLYL 318

DB 158

158 RTALDLADPSAKAVLTGEYKDELLESARSNGEKKMALLTPLNVNCHASDGRKSTPLHL 217

QY 319

319 AAVYGHLCDFRLLHGHADPDYNTDQGLLARVPRTTLEICLHNC---RPEYIQLLI 375

DB 218

218 AAGYNRVKIVQLLOHGA--DVHAKDKGDL--VP-----LHNACSYGHYEYVELLV 264

QY 376

376 DEGA 379

DB 265

265 KHGA 268

RESULT 7

US-09-196-387-10

Sequence 10, Application US/09196387

Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESS: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

THIS PAGE BLANK (USPTO)

Db 347 GGGGCGAGGTGGACCTGTGGATGTCAAGGGGAGAGCTGCCGTGATGGCTGTAGG 406
Qy 1105 CATGGCCATCTGAGCTGTGTAGCTGCTTTTGAAGCTGGCTCTCTCTGGGTAGC 1164
Db 407 AACGGGCACTTGGAGAGACGTAGATCTTTTGGAGGCTGGCTGATGCCAACGGCAGC 466
Qy 1165 ATCTCAACAACCTGTCTCCCGTGTCAAGCTGGCCGATGAGTGGCTGTATCCG 1224
Db 467 CGGACCCACCGCAGACCTCTGTGTATCAATGCTCTTGTGTGGTAGGAGCAGACATCTCG 526
Qy 1225 CAGAGCTCTAGACATAGTGTGCAAGGCGCAAGTCAAGCTAACT-----ACC 1274
Db 527 AAGGCTTATCAGGTATGGGGGAGATGTGTATCAACATCATCTGAATTTGACACCC 586
Qy 1275 AGTCTGGGATCAACATAGCTTCAATGTTCT-----GGCCCTCTATTGGCCGCA 1326
Db 587 CGGCCCCCTTTTTCAGGGGGCTTAACCTCTGTGGTGTGTCTCTTATCAATCAGTGT 646
Qy 1327 GTCTACGGGCACTGTGACTGTTTCCGCTGCTTTTGTCTCAAGGGGAGACCTGACTAC 1386
Db 647 GCTTACCATTAACCTTCAAGCTCTTCAAGGCTCTTGTGAGGCTGGGAGCAATCTGACTTC 706
Qy 1387 AACTGCACTGACCAAGGCG 1404
Db 707 AATTGCAATGGCCCTGTGC 724

RESULT 2

US-09-302-769-26
Sequence 26, Application US/09302769

Patent No. 6323317

GENERAL INFORMATION:

APPLICANT: HILTON, Douglas J

APPLICANT: ALEXANDER, Warren S

APPLICANT: VINEY, Elizabeth M

APPLICANT: WILSON, Tracey A

APPLICANT: RICHARDSON, Rachael T

APPLICANT: STARR, Robyn

APPLICANT: NICHOLSON, Sandra E

APPLICANT: METCALF, Donald

APPLICANT: NICOLA, Nicos A

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS

FILE REFERENCE: 109762

CURRENT FILING DATE: 1999-04-30

PRIOR FILING DATE: 08/962,560

NUMBER OF SEQ ID NOS: 50

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 419

TYPE: DNA

ORGANISM: Human

US-09-302-769-26

Query Match

Best Local Similarity 55.4%; Score 71.2; DB 4; Length 419;

Matches 164; Conservative 0; Mismatches 123; Indels 9; Gaps 1;

Qy 862 GAGAGACAGGCTCTCAATCAGAGTGTATGACAAGACTCTCTATCTTTGACAGCAGCTT 921
Db 114 GAGGACAGAGGCTCTCATATGACGTTACGTGGGAGCTTCAAGACCTTCAGAGCCTTA 173
Qy 922 TTGGCCAGAGACGTTCAACAGTTTCAATCAACAGAGAGTGGCTGG-----GGT 972
Db 174 TTGCAAGAGAGAGGCTACCGAGAGCCGATCAAGAAAGTGTGTGGTGTGGCTGG 233
Qy 973 GTTCTGGAGACCTTGGGCTGTGGCTGTCTTATGAGCACTTGAAGCTTTGCAAGTC 1032
Db 234 CTCCTTGCACACCGTTGGAAATCGGCGACCTGCAAGGCGCATGGAGCTGTGTGACCTTC 293
1033 CTCCTTAGCCCATGTGCTGATGTGACAGCTTGATGTCAAGGACAGAGCGCACTTTTC 1092

Db 294 CTATCCGGAAGGGGGCCGAGTGGATCTGTGGAGCTAAAGAGACAGAGCCCTGTAT 353
Qy 1093 ACTGCTGTCACTATGAGCCATCTGAGCTGTGTACGTGTCTTTTGAAGCTGTGC 1148
Db 354 GTGCTGTGTGTAACGGGCACTAGAGATACCAAGATCTCTTCGAAAGCTGGCGC 409

RESULT 3

US-09-221-298-28

Sequence 28, Application US/09221298

Patent No. 6284241

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS

FILE REFERENCE: 210121.471

CURRENT FILING DATE: 1998-12-23

NUMBER OF SEQ ID NOS: 112

SOFTWARE: PatsSeq for Windows Version 3.0

SEQ ID NO 28

LENGTH: 401

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: modified_base

LOCATION: (212)

OTHER INFORMATION: Where n is a, c, g or t

US-09-221-298-28

Query Match

Best Local Similarity 52.4%; Score 69.4; DB 3; Length 401;

Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

Qy 360 AAAAGAGTGGAGGAGGAGTGAATGGAATGATGATGATGATGATGATGATGATGAT 419
Db 52 AGAAGATTTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 111
Qy 420 GGAATATCCCAACAGAAATGAGACCATTAACATGATGATGATGATGATGATGATGAT 479
Db 112 GGGCTTCCCAATCAGCATGAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 171
Qy 480 CACCTACCTCCGGAATTAAGTGTCTTAATCTGTAATCTGTAATCTGTAATCTGTAATCT 539
Db 172 CACTTACCTGCTCTTTTGTGTTCCGTATGCTCTCANATATGATGATGATGATGATGAT 231
Qy 540 TTCAAGTTCAGAGTAAGAGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
Db 232 TGCAACTTTTATGATCCGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 291
Qy 600 TGCTGCAATTTGCGCTGTGAGCCAGCTCTCTGTGAT--TTACACTGCTGTGTA 656
Db 292 GACGCTATGTGTGCTGAGTCAAGCTCTTGTGCTATGATGATGATGATGATGATGATGAT 656
Qy 657 TGATGAGCTCTTGTGAGAGCCATTAGCCAAAGAA 692
Db 352 TGATGAGAAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 387

RESULT 4

US-09-155-078-1

Sequence 1, Application US/09155078

Patent No. 6312688

GENERAL INFORMATION:

APPLICANT: FOSTKA, Annemarie

APPLICANT: KIOSCHIS, Petra

APPLICANT: LAPORTE, Jocelyn

APPLICANT: HU, Ling Jia

APPLICANT: MANDEL, Jean Louis

TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN

FILE REFERENCE: 012627-010

CURRENT APPLICATION NUMBER: US/09/155,078

CURRENT FILING DATE: 1999-03-02
; EARLIER APPLICATION NUMBER: PCT/DE97/00592
; EARLIER FILING DATE: 1997-03-21
; EARLIER APPLICATION NUMBER: DE 196 11 234.6
; EARLIER FILING DATE: 1996-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3431
; TYPE: DNA
; ORGANISM: German-type microorganism & cell cul.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1863)
US-09-155-078-1

Query Match 3.6%; Score 67.2; DB 4; Length 3431;
Best Local Similarity 52.1%; Pred. No. 1.3e-09;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;
Qy 360 AAAAGATGAGGAAAGTGGATGGAAGTGAATGACCCCAATATCAGACTTTGGGCGTAT 419
Db 522 AGAAAGTTTAACTGGATGGATGACAGTTTTACAATCCAGTGAAGAATACAGAGGCA 581
Qy 420 GGAATACCCAAACAGAACTGACCATACAGATGCCCAACAGAACTATGAGATATGAG 479
Db 582 GGGCTGCCCATCACCATTGGAGAAATATTTTAAATAGTCTATGAGCTCTGTGA 641
Qy 480 CACCTACCCCTCCTGAAATAGTGGTTCTTAAATCTGTACCTTGGGAACGGTGGTGAAG 539
Db 642 CACTTACCCCTCTCTTTTGGTGGTTCCGTATCGTCCCTCAGATGATGACCTCCGGAGAT 701
Qy 540 TTCAAGTTTCAAGATGAAGACGTGTCCTGCTCTCTACCTCTTCAACAGAGAACAA 599
Db 702 TGCACCTTTTAGTCCCGAAATCGAATTCAGTGTGTCTATGAGTTCATCCAGAAATAA 761
Qy 600 TGCTGCCATTGCGCTGTAGCCGCTCTCTCTCGATT---TTACACTCGCTGTAGA 656
Db 762 GACGGTATTGCGTGTGATGACGCTCTTGTGCGGTATGAGTGGGAACGAAATAAAGA 821
Qy 657 TGATGAGTCTCTTGTGGAGGCGCATTAGCCAAACAAA 692
Db 822 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 857

RESULT 5
US-08-181-629A-2
; Sequence 2, Application US/08181629A
; Patent No. 5472872
; GENERAL INFORMATION:
; APPLICANT: Swaminathan, Neela
; APPLICANT: Van Ethen, James
; APPLICANT: Mead, David
; APPLICANT: Skowron, Piotr
; TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,629A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-181-629A-2

Query Match 3.5%; Score 66.6; DB 1; Length 5496;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 144; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
Qy 978 TGGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGGCTGTGGCAAGTCCCTTT 1037
Db 3468 TTGGACGGCGCTTACATTTACCGCGCTTTTAAATGGTATAGCATGTGCGTCAAGACGCTAT 3527
Qy 1038 AGCCATGGTGGCTGATCTTGACAGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGC 1097
Db 3528 TGATGGGGTGGCAATCTTGACATACAGATATTTCCGGATGTACACCATTTTCATCGTGC 3587
Qy 1098 TGTCACTGATGGCCATCTGGACTGTGTACGTGTGCTTTTGGAGCTGGTCTCTCTGG 1157
Db 3588 GGTATTAATGACCAAGATGATGTGTGAAGACTCTGTAGAGCAGTGCAACTTTGA 3647
Qy 1158 TGTAGCATCTACAACTGTCTCCCGTGTCTCAGCTGCCGCGTATGGTGTCTTGC 1217
Db 3648 CGTCAITGATGATCTGAGTGGGCGCTTACATTAACGCGCTTTTAAATGTAATGATGC 3707
Qy 1218 TATCTCGAGAGCTCTAGACCATGCTGTCAGA 1250
Db 3708 GATTTGAGATGCTCATTAAGCAGGTGCAGA 3740

RESULT 6
US-09-509-802-1
; Sequence 1, Application US/09509802
; Patent No. 6489130
; GENERAL INFORMATION:
; APPLICANT: Immunex Corp.
; APPLICANT: Bird, Timothy
; APPLICANT: Virca, G.D.
; TITLE OF INVENTION: DEATH ASSOCIATED KINASE CONTAINING ANKYRIN REPEATS (DAKAR)
; FILE REFERENCE: 2889-US
; CURRENT APPLICATION NUMBER: US/09/509,802
; CURRENT FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 1
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-509-802-1

Query Match 3.2%; Score 60.2; DB 4; Length 2370;
Best Local Similarity 51.3%; Pred. No. 1.2e-07;
Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
Qy 980 GGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGGCTGTGGCAAGTCCCTTTAG 1039
Db 1829 GGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGGCTGTGGCTGGCTTCCATTTCAT 1888
Qy 1040 CCATGCTGTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGCTG 1099
Db 1889 ACCTGCTCTGTATGTTAATCATCTGACGCTTACAGCCTTACAGACACACCTTGCATGTTGCTG 1948